Describing the evolution in the number of highway deaths by a decomposition in exposure, accident risk and fatal risk

Peer-reviewed author version

HERMANS, Elke; WETS, Geert & VAN DEN BOSSCHE, Filip (2006) Describing the evolution in the number of highway deaths by a decomposition in exposure, accident risk and fatal risk. In: Tansport Research Record. p. 1-8.

Handle: http://hdl.handle.net/1942/1364

Describing the Evolution in the Number of Highway Deaths by a Decomposition in Exposure, Accident Risk and Fatal risk

Elke Hermans, Geert Wets* and Filip Van den Bossche

Hasselt University
Transportation Research Institute
Agoralaan, Building D
3590 Diepenbeek
Belgium
Fax: +32(0)11 26 87 00
Tel: +32(0)11 26{87 76, 86 49, 87 76}
E-mail: {elke.hermans; geert.wets; filip.vandenbossche}@uhasselt.be

Total number of words: 7346 (tables and figures included)
Resubmission date: 15/11/2005

* Corresponding author

ABSTRACT

The general purpose of this research is to improve the insight in road safety on Belgian highways by means of a layered model. In this study, the monthly number of persons killed on highways in Belgium will be decomposed in three parts, namely exposure, accident risk and fatal risk. The evolution in these three dimensions will be investigated separately. More specifically, for each dimension a descriptive and explanatory analysis will reveal the most optimal unobserved components model. The separate analysis of each dimension may reveal different underlying developments. We will study the impact of meteorological, socio-economic, legislative and calendar factors on exposure, accident risk and fatal risk. The analysis shows that, although for each dimension the same basic components are available, the optimal model of each dimension has its unique structure of descriptive components and significant variables. Precipitation and snow enhance accident risk while temperature plays a significant role for exposure. Fatal highway risk decreases in case of an extra day with precipitation and was significantly affected by the child restraint law. The economic indicators mainly affected accident risk. Bringing the three models back together shows that the number of highway deaths during 1993-2001 has well been reconstructed.

INTRODUCTION

During the past decennia there has been a steady increase in traffic volume, which resulted in continuously increasing traffic problems. Worldwide, an estimated 1.2 million people are killed in road crashes each year and as many as 50 million are injured (*1*). In the battle for road safety, several countries have set themselves an ambitious goal to reduce the number of people killed between 2000 and 2010 by half (*2*). The increasing interest in a traffic safe world is an incentive for the elaboration of numerous kinds of models. An important family of road safety models is based on time series analysis. A time series consists of the repeated measurement of a certain phenomenon at regular intervals. In that way, the evolution of road safety in time can be studied. In general, models are used to describe the data, to explain the time series and to forecast. The first two aspects will be handled in this study.

The two central aspects in this study are road safety dimensions on the one hand and time series components on the other. Each road safety problem can be decomposed into three dimensions: exposure, accident risk and loss (*3)*:

$$\boxed{\begin{aligned}&\text{Road Safety} = \text{Exposure} \times \text{Accident Risk} \times \text{Loss} \;\text{ or}\\ &\text{Road Safety} = \text{Exposure} \times \frac{\text{Number of Accidents}}{\text{Exposure}} \times \frac{\text{Number of Victims}}{\text{Number of Accidents}}\end{aligned}}$$

Exposure is the amount of movement or travel within a traffic system. It can be measured in several ways, for example by counting traffic, travel habit surveys and indirect exposure estimates. In this study,the exposure unit is the number of vehicles counted on highways. The second dimension, accident risk, is the underlying probability of a crash, given a particular exposure. The third dimension is the probability of injury in case of a crash. The three dimensions that will be used are exposure, accident risk and fatal risk. The last one is defined as the ratio of the number of persons killed and the number of accidents.

A road casualty results from the presence of a person on the road, the fact that the person got involved in an accident and got injured. Such a decomposition enhances our knowledge since more information about the underlying aspects of an accident becomes available. The different level of road safety in several countries can be explained by these three dimensions. Additionally, a study of the evolution in these dimensions over time will possibly yield valuable insights which are impossible to detect at the aggregate level. The investigation of each dimension reveals underlying risk factors. Although there exists general agreement on the validation of this formula, little is known about the specific factors affecting each dimension. Road safety decision making can benefit from this knowledge since particular measures are needed in order to reduce accident risk or loss or to influence exposure.

There has already been done research regarding decompositions of the number of accidents or fatalities. In 1948 Smeed (*4*) compared the total number of fatalities (D) in 20 countries to the number of cars (N) and the size of the population (P) in each country by means of a regression model. Smeed came to his famous formula $D = 0.0003(NP^2)^{1/3}$. When at the beginning of the seventies, the number of fatalities decreased in almost all Western countries – a phenomenon contrary to Smeed's expectations – this formula became invalid.

The progress of time series analysis techniques as well as the availability of more detailed time series data have led to advanced and interesting statistical studies on road traffic safety (*5*). In 1982 Appel (*6*) found an exponentially decaying risk when he decomposed the (expected) number of accidents in a risk component and exposure. Similar approaches have been adopted by Broughton (*7*) and Oppe (*8*).

One special and prominent class of explanatory models is known as the DRAG (Demand Routière, les Accidents et leur Gravité) family, extensively described in (*9*). A DRAG analysis consists of three independently modeled stages – traffic volume, risk expressed in accidents per kilometer and the number of victims per accident. DRAG models are structural explanatory models, including a relatively large number of explanatory variables, whose partial effects on the exposure, the frequency and the severity of accidents are estimated by means of ARIMA(X) (Auto-Regressive Integrated Moving Average with eXplanatory variables) techniques.

Recently, Bijleveld and Commandeur proposed the Basic Evaluation Model (BEM) which contains features of the Smeed, Oppe and DRAG approaches while also allowing for more flexibility and a more intuitive interpretation of results (*10*). The BEM is a bivariate state space model which models mobility with a trend component that also appears in the equation for fatal accidents beside its own trend. Multiplying both trends approaches the number of fatal accidents.

This study is different from previous research as the underlying dimensions of the road safety problem are modeled using the unobserved components methodology (UCM). We will use a layered structure – exposure, accident risk and fatal risk – but instead of ARIMA(X) techniques (used in DRAG models) we opt for the unobserved components methodology. Different from the current BEM is the testing of several explanatory variables.

The starting point for the construction of unobserved components models is the representation of a (log-transformed) observed value as the sum of a trend, a seasonal and an irregular component.

$$\boxed{y_t = \mu_t + \gamma_t + \varepsilon_t \text{ for } t = 1,\ldots,n}$$

For example, the logarithm of the observed number of accidents (y_t) in January 2000 can be divided into the following three parts: part 1 represents the long-term trend (μ_t) – consisting of a level and slope – which has been slowly decreasing every year, part 2 takes into account the fact that this is a January month and is a periodic (or seasonal) component (γ_t) and part 3 contains some unique facts for that specific month (ε_t). Each component is modeled by a separate dynamic process. Beside the descriptive components – level, slope and seasonal – the model can be extended by adding explanatory variables and intervention effects (*11*).

This study attempts to find the optimal model for each of the three dimensions. In the descriptive analysis based on level, slope and seasonal we try to obtain the model that fits the data best. It is assumable that highway exposure, accident risk and fatal risk have a unique descriptive structure. Furthermore, studying the impact of meteorological, socio-economic, legislative and calendar factors on the three dimensions will teach us more than an aggregate analysis can. When for example the effect of precipitation has an opposite sign for accident and fatal risk, no significant impact will be noticeable at the aggregate level although precipitation can have an important effect on these dimensions. In the end, the three optimal models will be brought back together and compared with the number of deaths.

The remainder of this text is organized as follows. First, we give a short introduction to the unobserved components methodology. Next, the data will be discussed. The method of working and results will be described in the following section and this paper closes with conclusions and topics for further research.

METHODOLOGY

As mentioned before, the unobserved components methodology (*12*) is central in this study. UCM models distinguish themselves on a number of aspects. They provide an explanation of the main features of the phenomena under investigation. Models are set up explicitly in terms of the components of interest, such as trends and seasonals. In addition, instead of assuming that these components remain constant over time, this approach allows them to evolve. UCM is often used in different research fields. Beside economic applications (*13*,*14*), this technique and more specifically intervention analysis was also used in traffic related research (*15*). The methodology forms a well-used approach in modeling road accidents in several countries (*16*,*17*).

The key to handling UCM models is the state space form, with the state of the system representing the various unobserved components (*12*). Once in state space form, the Kalman filter (*18*) may be applied and this in turn leads to estimation, analysis and forecasting. The overall objective of the state space analysis is to study the development of the state over time using observed values (*19*). More specifically, we want to obtain an adequate description and find explanations for this development. A state space model consists of an observation or measurement equation, state equations and an equation which specifies the initial condition of the state elements. The observation equation contains the unobserved state at time t and a white noise observation error. In the state equation, time dependencies in the observed time series are dealt with by letting the state at time t+1 be a direct function of the state at time t and the state error is also white noise. Algebraically, a complete state space model can be written as:

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \sum_{i=1}^{l} \lambda_{it} w_{it} + \varepsilon_t \qquad \varepsilon_t \sim NID(0,\sigma_\varepsilon^2) \qquad (1)$$

$$\mu_{t+1} = \mu_t + v_t + \eta_t \qquad \eta_t \sim NID(0,\sigma_\eta^2) \qquad (2)$$

$$v_{t+1} = v_t + \zeta_t \qquad \zeta_t \sim NID(0,\sigma_\zeta^2) \qquad (3)$$

$$\gamma_{t+1} = -\sum_{m=1}^{s-1} \gamma_{t+1-m} + \omega_t \qquad \omega_t \sim NID(0,\sigma_\omega^2) \qquad (4)$$

$$\beta_{j,t+1} = \beta_{jt} + \tau_{jt} \qquad \tau_{jt} \sim NID(0,\sigma_\tau^2) \qquad (5)$$

$$\lambda_{i,t+1} = \lambda_{it} + \xi_{it} \qquad \xi_{it} \sim NID(0,\sigma_\xi^2) \qquad (6)$$

for $t = 1,\ldots,n\,;\ j = 1,\ldots,k\,;\ i = 1,\ldots,l;\ m = 1,\ldots,s-1$

The observation equation (eq. 1) relates the values of the dependent variable y_t to the trend μ_t, the seasonal component γ_t, explanatory variables x_{jt} and intervention variables w_{it}. Each component has its state equation (eq. 2-6 respectively). Note that the slope component (v_t) appears also in the state equation of the trend. There are various ways in which a seasonal can be incorporated (*20*). In equation 4, a dummy seasonal is presented. In state space methods, time series data are assumed to be stochastic, thus errors are included in all equations (*21*). All errors are assumed to be mutually independent and normally distributed. β_j is the estimated regression coefficient of the j^{th} explanatory variable. One type of intervention is the temporal impulse-intervention. In that case, w_{it} takes value 1 if t is the month which needs correction for a special event. λ_i is the coefficient of the i^{th} correction variable. The parameter estimations are obtained with an iterative process, using the maximum likelihood principle. The software used in this research is SsfPack (*22*), a suite of C routines linked with the Ox matrix programming language (*23*).

Compared to other models, these models gain in flexibility because the stochastic formulation allows the components to evolve over time. However, we try to obtain a model as parsimonious as possible and opt for one (fixed or deterministic) parameter estimation during the entire time period when this is a good approximation. In that case, the error term is zero and the corresponding state equation becomes redundant. In case a certain component does not contribute to the description or explanation of the model, this component can be omitted from the state without loss of information but again gain in efficiency. The above model is thus the general, complete model which can be adjusted in order to obtain the best possible fit with the least parameters.

DATA

Since we want to gain insight into the evolution of road safety over time, a large number of observations is necessary. Monthly observations are often preferred as a good balance between data availability and data variability. The data used are monthly observations from 1993 till 2001. The two key data sources we need are a measure of exposure on the one hand and accident related data on the other hand. We decided to study the road safety on Belgian highways since highway exposure data are available. The number of vehicles on Belgian highways (per thousands) are known for each month during the specified 9 years. Because we choose to make a decomposition into three parts, beside the number of highway accidents, we need data about the number of victims. In this perspective we opt for the number of deaths (any person who was killed outright or died within 30 days as a result of the accident). We work with the most reliable data we can find. They all have been gathered from governmental ministries and official documents published by the Belgian National Institute for Statistics.

As the first graph in Figure 1 shows, the exposure observations (HEXPOS) are peculiar. At the beginning of 1999 the counting method changed. By inserting a level intervention in our model for January 1999, the increase will not be treated as part of the actual evolution in exposure but as an exceptional fact. The two accident related outcomes are presented in graphs 2 and 3. The number of highway accidents (NHACC) increased over the years while the number of persons killed on a highway (NHKIL) remained constant during the decade. The graphs also give an indication of the existence of differences between months. Furthermore, the more than average high or low values during some time points attract attention.

The vehicle counts presented in graph 1 of Figure 1 are used as an indicator of exposure. The second dimension – highway accident risk (HACCRISK) – is presented in graph 4. It is the ratio of the number of accidents on highways and highway exposure. The third dimension (graph 5) is the risk of being killed in a highway accident (HKILRISK) and resulted from the division of the number of persons killed in a highway accident and the number of highway accidents. The analyses will show to what extent the trend in the number of persons killed depends on a change in the number of vehicles on the road, another accident risk and/or a change in the probability of being killed in a highway accident.

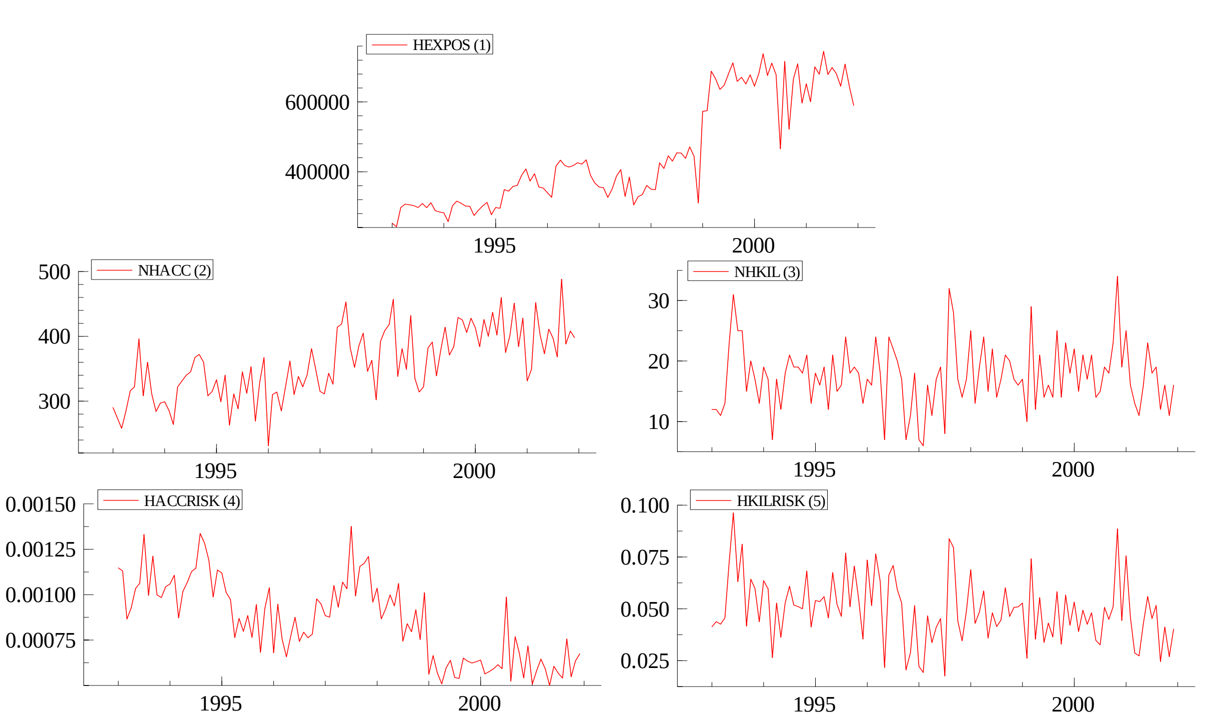

FIGURE 1 Actual monthly observations of the number of counted vehicles on Belgian highways, the number of highway accidents, the number of persons killed on a highway, the highway accident risk and the risk of being killed in a highway accident from 1993 till 2001.

Our objective is to further investigate the three dimensions and more specifically to explain the long term trend. Therefore, extra information is needed. As in most areas of the social sciences, there is no firm theory indicating which explanatory variables should appear in a model (*24*). The choice of independent variables must be guided by research and professional judgment. In this study, the effect of 9 explanatory variables, related to climate, economy, laws, demography and time, proven important in earlier research (*16*,*24*,*25*,*26*), will be investigated. In order to reduce multicollinearity to an acceptable level, several variables were excluded from the original database counting 23 explanatory factors. Based on the variance inflation factors (*27*) we performed a stepwise elimination of the variables with an unacceptable effect on the variances. Although it is regrettable that the impact of variables like frost, average fuel price and the regulation concerning hands-free calling can not be tested in this study, at least we are convinced that the obtained regression estimates will be unbiased. Although a fixed database of 9 explanatory variables is used for all dimensions, we may assume that different factors will be significant for the three dimensions.

The explanatory variables of the first group represent climatic conditions concerning precipitation (NDAYSPREC), snow (NDAYSSNOW) and temperature (AVGTEMP). These data have been collected by the Belgian Royal Meteorological Centre. The temperature is expected to particularly influence exposure, while rain and snow will rather have an impact on the risk dimensions. Given the fact that snow falls on average only during 19 days a year, the theory of risk compensation (*28*) is possibly valid. The next three variables are all economically related. With the percentage of unemployed persons (PERUNEMP), the number of car registrations (NCAR) and the monthly indicator of the Belgian National Bank (NBBECTRD) we want to investigate the effect of the economy on the dimensions underlying the number of persons killed. We will also study the impact of regulations. Due to multicollinearity we test only one law which was imposed in Belgium during the period 1993-2001 and furthermore of which we could expect an effect on the road safety on highways. In September 1996 a regulation concerning child restraints was imposed (LAW0996). We do not expect a significant impact on exposure and accident risk. After all, this law has been imposed to reduce the severity of injury in case of an accident. Next, we brought one demographic variable in the analysis, namely the natural population growth (NATPOPGROWTH) in Belgium. Other demographic variables worth investigating – like the distribution over age classes and sex – are only available on a yearly basis. Monthly information would be valuable since each demographic group has its own travel and risk pattern (*29*). Finally, we test the calendar factor ‘number of weekend days and holidays’ (NWEHOL) of which we particularly expect an impact on exposure.

METHOD

The objective of the descriptive analysis is to select out of several models the one that describes the evolution in the data best. We study exposure, accident risk and fatal risk separately. As mentioned before, there are three descriptive components – level, slope and seasonal – which can be chosen stochastically or deterministically. In order to know which components must be included in the model we look at their contribution. If the variance of the error term in the state equation is (close to) zero, the component is assumed to be deterministic. A component must be included in the model if it is significant. We work with a confidence interval of at least 90%. When we follow this reasoning, the resulting model will probably have the lowest Akaike Information Criterion (AIC), which will be used here as an indicator of fit (*30*). Nevertheless, an AIC value is only important if the three error terms conditions – no serial correlation (Q), homoscedasticity (H) and normality (N) – are fulfilled. The diagnostic tests we use are respectively the Ljung-Box statistic (*31*), the H-test (*15*) and the Bowman-Shenton test (*32*). Adding variables or applying data transformations can be helpful to meet these assumptions. A particular concern is the existence of outliers and breaks in a time series since they can distort the estimation of parameters and are too influential in the empirical analysis (*5*). In the context of UCM, one can study the auxiliary residuals. A relatively large observation error indicates the presence of an outlier while a relatively large value in the level noise indicates a structural break.

We start with the first dimension, highway exposure. The first graph in Figure 1 showed that a level intervention in January 1999 was necessary to take into account the increase as from January 1999 due to a change in counting method. For the first model we add to that a stochastic level, slope and seasonal. Based on the variance of the error terms in the state equations, the *t*-values of the components and the auxiliary residuals, we derive the best model. This contains a stochastic level, a deterministic slope, a deterministic seasonal, a level intervention (Jan99) and four irregular interventions (Mar97, Dec98, Jul00 and Sept00).

Although we can not come up with an explanation for the extreme values during four time points, we are convinced that the most striking shocks must be excluded in order to fulfill the error terms conditions. In the end, we want to obtain the optimal model and a correct parameter interpretation. Although the above specified model is the best one, the homoscedasticity value of 3 is far above the critical value of 1.77, thus unacceptable. A possible solution is to transform the original data, for example by a logarithmic transformation (LHEXPOS). The problem of heteroscedasticity was resolved. The consequence however is that for the other two dimensions we must also work with log-transformed data. The decomposition of the number of persons killed changes then from a multiplication into an addition of log-transformed dimensions:

$$\boxed{\mathrm{LNHKIL}_t = \mathrm{LHEXPOS}_t + \mathrm{LHACCRISK}_t + \mathrm{LHKILRISK}_t}$$

We repeat the working method for LHEXPOS and the same model turns out to be optimal. The most important characteristics of this best descriptive model are given in Table 1. Empty cells represent insignificant or irrelevant combinations. For the remainder of the study we will work with log-transformed dependent variables.

For the two remaining dimensions the same procedure is applied. Since we may expect that the change in counting method influences the accident risk, we also test the inclusion of a level intervention in January 1999 for the second dimension. The variances, *t*-values and residuals make us come to a model with stochastic level, no slope, deterministic seasonal, level intervention in January 1999 and irregular intervention in July 2000. The lowest AIC value is obtained for this model. Characteristics of the model are reported in Table 1.

The results of the model with three stochastic components for the logarithm of the fatal highway risk (LHKILRISK) show that a model with deterministic level and slope yields the best fit. The logarithmic transformation causes the normality to exceed the critical value of 5.99. Since the residual graph does not indicate that for certain time points outliers must be incorporated in the model, we keep this model. More information can be found in Table 1.

The second objective of this study is to find an explanation for the evolution over time of the three dimensions. 9 independent variables will be tested here. First, the entire set of explanatory variables will be added to each best descriptive model. We choose a confidence interval of 90%, therefore a *t*-value in absolute value higher than 1.645 guarantees the inclusion of the variable into the model. The AIC will be used to compare the fit between models. The second explanatory model for each dimension retains only the significant variables in order to improve the parsimony of the model.

The analyses showed the following. For highway exposure 2 of the 9 variables seemed significant, namely the average temperature and to a lesser extent the indicator of the National Bank. Most explanatory variables were significant on the accident risk dimension. The number of days with precipitation, the number of days with snow, the number of car registrations and the National Bank indicator added to the best descriptive model of LHACCRISK yields an AIC of -4.197. For LHKILRISK the number of days with precipitation seemed significant, as well as the law concerning child restraints. Adding explanatory variables resulted in a no longer significant slope, which we excluded from the final model. The characteristics of the final model (significant explanatory factors included) for each dimension are given in the lower part of Table 1.

TABLE 1 Model Characteristics of the Best Descriptive and Final Model for the Three Dimensions

	LHEXPOS	LHACCRISK	LHKILRISK
BEST DESCRIPTIVE MODEL			
Descriptive statistics			
AIC	-5.5844	-3.9874	-2.2038
Q(9,8) < 15.51	8.4829	9.1188	11.593
H(35) < 1.77	1.1229	0.6299	1.2835
N(2) < 5.99	5.3651	0.4038	7.2231
t-values descriptive components			
level	180.07	-59.30	-51.00
slope	1.64		-2.60
level outlier (Jan99)	4.74	-3.62	
irregular outlier (Mar97)	-3.68		
irregular outlier (Dec98)	-5.33		
irregular outlier (Jul00)	-7.75	3.82	
irregular outlier (Sep00)	-4.60		
	LHEXPOS	**LHACCRISK**	**LHKILRISK**
FINAL MODEL			
Descriptive statistics			
AIC	-5.6984	-4.1969	-2.2826
Q(9,8) < 15.51	6.6984	7.6824	9.4232
H(35) < 1.77	1.3238	0.6748	1.4510
N(2) < 5.99	4.6400	0.6912	6.4854
t-values descriptive components			
level	168.99	-35.51	-27.40
slope	1.86		
level outlier (Jan99)	4.26	-4.14	
irregular outlier (Mar97)	-4.44		
irregular outlier (Dec98)	-6.30		
irregular outlier (Jul00)	-7.49	3.91	
irregular outlier (Sep00)	-5.00		
Coefficients explanatory variables (t-value)			
NDAYSPREC		0.0051 (2.37)	-0.0112 (-1.99)
NDAYSSNOW		0.0134 (2.62)	
AVGTEMP	0.0098 (3.43)		
NCAR		-3.8e-006 (-2.36)	
NBBECTRD	-0.0029 (-1.65)	0.0061 (2.02)	
PERUNEMP			
LAW0996			-0.2138 (-3.45)
NATPOPGROWTH			
NWEHOL			

RESULTS AND DISCUSSION

Descriptive Analysis

The four graphs in Figure 2 recapitulate the descriptive analysis. Graph 1 shows the trend of the model with stochastic level, deterministic slope, deterministic seasonal and 5 outliers for the log-transformed exposure on Belgian highways. The second graph shows the trend of the model with stochastic level, deterministic seasonal and two outliers for the log-transformed risk of having an accident on a Belgian highway. The third graph visualizes the trend estimation of the deterministic level and deterministic slope model for the log-transformed risk of being killed in a highway accident. The trend line in the fourth graph has been realized by summing up the three previous trends for all time periods. The trend in the number of persons killed on a Belgian highway has remained more or less constant from 1993 till 2001. It is however interesting to be able to look at the evolution in the underlying dimensions. Generally, there has been an increase in the number of vehicles on Belgian highways, which could intuitively be expected. The risk of having an accident on the highway has decreased in a fluctuating way. The fatal risk trend is linearly decreasing. Given the fact that an accident on a highway happens, the probability to die does not depend (significantly) on the month. Over time this probability has come down due to amongst other things quicker interventions, progress in medical treatment and better car technology.

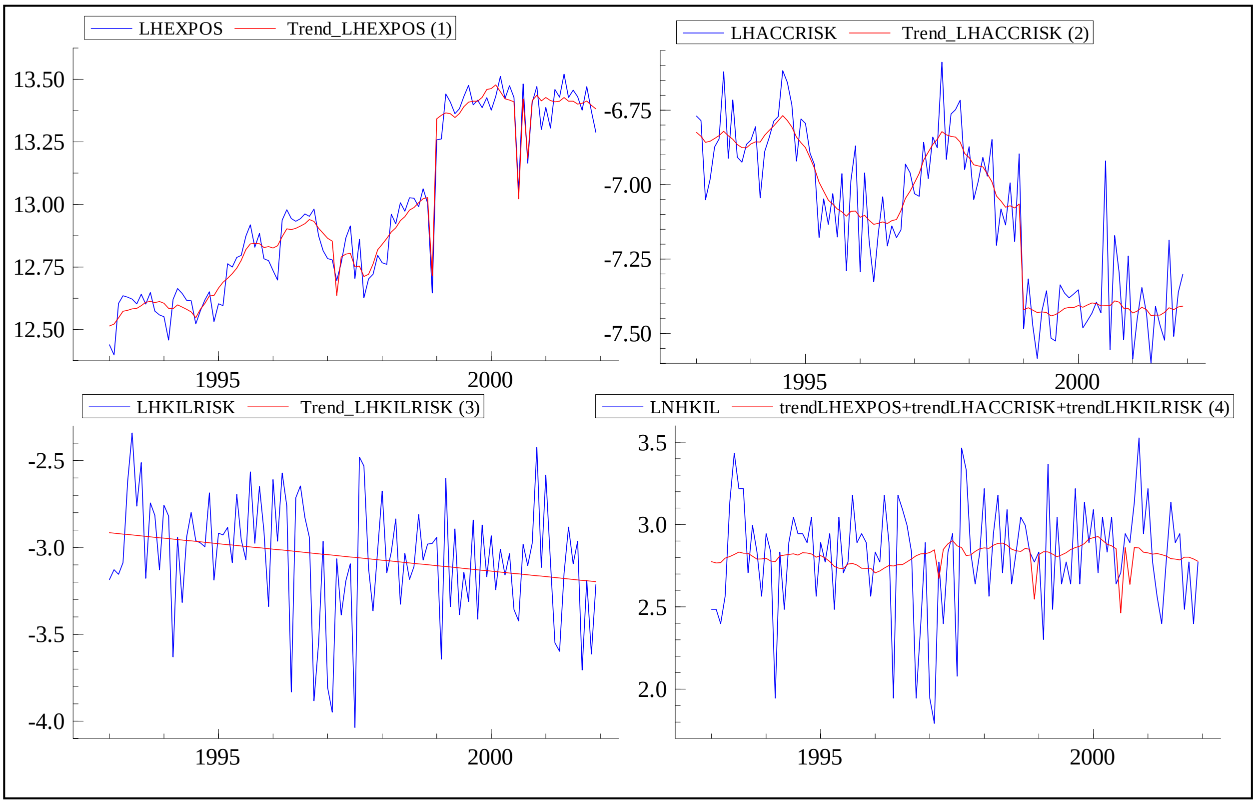

FIGURE 2 Actual monthly observations and trend estimation of the best descriptive model for the three dimensions and construction of the trend of the number of highway deaths.

Explanatory Analysis

The explanatory analysis showed that the evolution in the trends of the three dimensions can partially be explained by meteorological, economic and legislative variables (see Table 1).

Weather conditions have a significant impact on road safety. We observe however that the effect is strongly related to the geographical properties of the area of concern and the level of aggregation (*33*). In this paper we tested the effect of snow, precipitation and average temperature. The findings of Zhang et al. (*34*) showed that the accident risk under adverse weather is higher than under non-precipitation weather and the highest risk occurs under snow conditions. Fridstrøm et al. (*24*) also found that rainfall is liable to increase the accident toll. Our study concludes that the monthly number of days with precipitation influences both highway accident and fatal risk. The signs tell us that a higher monthly percentage of days with precipitation causes for a given exposure, more accidents but less fatal accidents. A slippery surface, lower visibility, the treacherous feeling of getting used to constant rainy weather and need for more concentration are possible reasons for the higher accident risk. The “negative” impact on fatal risk can partly be explained by a lower average speed in wet conditions and consequently less serious accidents. These results are in line with Eisenberg (*35*) who found that compared to fatal crash rates, non-fatal crash rates are increased more by a given amount of precipitation. Another research (*24*) concluded that in case of snowfall and frost drivers may be under the risk compensation hypothesis and adjust their driving habits so as to more or less offset the increased hazard due to slippery road surfaces. This study found a positive relationship between snow and accident risk. A reduced exposure during winter conditions could not be concluded from our study but in general highways are a type of road from which snow is cleared away primarily. Third, the results show that exposure is significantly higher in case of an increase in average temperature. A rise of one degree in the average temperature increases exposure by exp(0.0098)-1 or 0.98%.

Previous research has suggested that economic activity is at least a partial explanation for changes in road safety. In our study, we tested the influence of three economic factors. Although the impact of unemployment proved significant at the aggregate level in other studies, the percentage of unemployed persons did not play an explanatory role in this study. Furthermore, no economic factor had a significant impact on fatal risk. Although the number of car registrations intuitively affects exposure, only a negative significant effect on highway accident risk was found. During the last years, the share of second-hand car registrations in the total number has decreased every year (*36*). An increase in the number of new cars implies a higher quality level of the vehicle fleet. The National Bank indicator showed that a better economic climate decreases the number of vehicles counted on the highway (this slight significant effect is counterintuitive) and increases accident risk, ceteris paribus. An explanation of this can not be found.

The analysis of the effects of known events such as a political decision is called intervention analysis (*37*). In earlier decades, an established set of interventions – related to speed, alcohol, protective devices, design of roads and vehicles – has contributed to significant reductions in the incidence and impact of road traffic injuries (*1*). We distinguish two possible intervention effects, namely a pulse intervention and a step intervention (*21*). The first effect is used to capture single special events since they may cause outlying observations which the pulse regression variable takes outside the general model. In the descriptive analysis we added some outliers of this type. The second intervention – called a step intervention or level shift – is introduced in the model to capture events such as the introduction of new policy measures. In this study, the impact of one law was investigated. The regulation concerning child restraints proved significant for the fatal risk on Belgian highways and its introduction resulted in a 19,25% (=exp(-0.2138)-1) reduction in fatal highway risk. We remark however that in our model it is assumed that the introduction of a law results in a sudden and permanent decrease in the dependent variable and that this assumption is not always a natural one.

Reunion of the Dimensions

Beside studying the evolution in the trend, we could try to reconstruct the number of persons killed on Belgian highways by adding the model estimates of the three dimensions. Because the optimal model for the fatal risk dimension does not contain a seasonal component, the reconstruction levels off and does not reach the high and low points of the actual observations (see upper graph in Figure 3). In the lower graph of Figure 3 the estimated values of the first two dimensions are summed in order to reconstruct the number of highway accidents. It shows how well the actual observations are approached by the decomposed model.

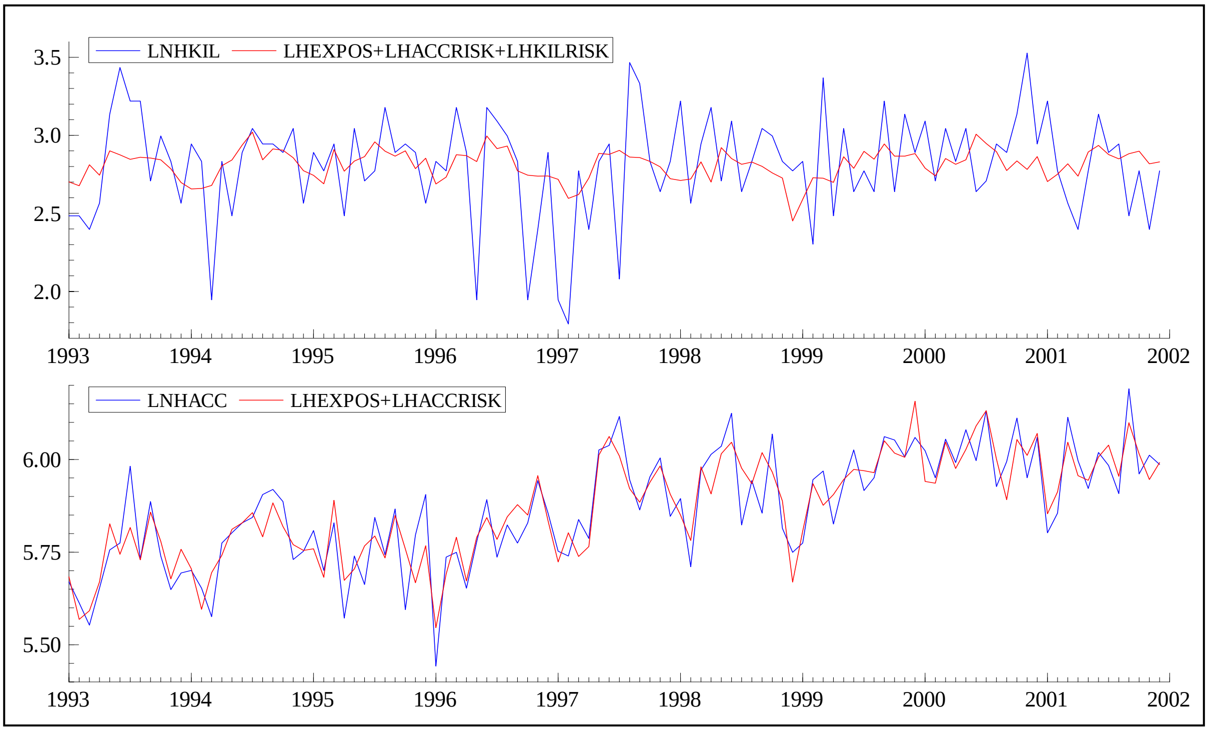

FIGURE 3 Reconstruction of the number of persons killed on a highway and the number of highway accidents based on the final models.

Implications

During the period 1993-2001 the number of deaths on highways in Belgium has remained more or less constant. The increase in exposure has been counterbalanced by a reduction in both accident and fatal risk. In order to reach the objectives set out for the future, still a lot of effort has to be made. The layered analysis in this study gave us new insights in the specific influences of variables. The following policy recommendations can be formulated: the monthly number of days with rain even as the monthly number of days with snow both increase accident risk. By means of dynamic road management or information campaigns drivers must become aware of the fact that given a certain number of vehicles on the highway the risk of getting involved in an accident is significantly high in case of rain and even higher in case of snow. However, given an accident in rainy conditions, the outcome is less severe because drivers probably adjust their average speed. Next, exposure seems to be most affected by the average monthly temperature. Unfortunately, no demographic or time effect could be found. Further, the introduction of a law obligating the use of child restraints was successful and quantified.

CONCLUSIONS AND FURTHER RESEARCH

This study showed the advantages of decomposing a road safety problem. The evolution in three dimensions – exposure, accident risk and fatal risk – has been described and explained separately. The unobserved components methodology has been proven useful for decomposition analysis. The basic components for all dimensions were the same, namely a level, slope, seasonal and 9 explanatory variables which could be included stochastically or deterministically. The final, optimal unobserved components model for each dimension however had its own, unique structure. The best model without explanatory variables described the evolution in the trend already quite well. This could have influenced the relatively small number of significant explanatory factors. Significant impacts could be noted for three climatic variables, two economic factors and the law concerning child restraints. Most of the tested variables seemed significant for accident risk.

For Belgium exposure data are rarely available on a monthly basis. Only vehicle counts on highways during several years were on hand, which restricted this research to road safety on highways. The exposure data however are far from perfect. In the future, artificial data adjustment is an option in order to reform the observations to more realistic values. Also a remark concerning the time period can be made. Nine years of observations limits the analysis, especially because the latest known accident figures date back to December 2001. As soon as more actual data become available the predictive power of our models can be tested and a forecast for the future development in all trends can be made.

Although a number of explanatory variables has been tested in this study, we can still think of several factors of which the impact would probably differ across the dimensions. The percentage of motors, cars and trucks as well as the size of the vehicle park would be worth investigating. Therefore, an acceptable tool must be found to transform data known on a yearly basis to monthly values. In the future, it could also be worthwhile to select per dimension a number of (uncorrelated) variables which are expected to have an effect and work with a specific explanatory dataset for each dimension.

Another point of further research is more model related. In this study the first attempt of developing a layered structure has been described. In a next phase, the dimensions should no longer be handled independently. Exposure should be included in the model equation of accident risk, which at its turn, should appear in the model equation of fatal risk. This multivariate model would be more complex but at the same time more realistic.

REFERENCES

(1) World Health Organization. World Report on Road Traffic Injury Prevention. www.who.int/world-health-day/2004/infomaterials/ world_report/en. Accessed March 20, 2004.

(2) European Commission. *White Paper "European Transport Policy for 2010: Time to Decide"*. Publication COM(2001)0370. Commission of the European Communities, 2001.

(3) European Commission. *COST Action 329: Models for Traffic and Safety Development and Interventions.* Publication EUR20913. Commission of the European Communities, 2004.

(4) Smeed, R. J. Some Statistical Aspects of Road Safety Research. *Journal of the Royal Statistical Society A*, Vol. 112, No. 1, 1949, pp. 1-34.

(5) Bijleveld, F. D., J. J. F. Commandeur, S. J. Koopman and K. van Montfort. Nonlinear Time Series Aspects of Road Safety Research. 2004.

(6) Appel, H. Strategische Aspekten zur Erhöhung der Sicherheit im Straßenverkehr. *Automobil-Industrie*, Vol. 3, 1982, pp. 347-356.

(7) Broughton, J. Forecasting Road Accident Casualties in Great Britain. *Accident Analysis and Prevention*, Vol. 23, No. 5, 1991, pp. 353-362.

(8) Oppe, S. The Development of Traffic and Traffic Safety in Six Developed Countries. *Accident Analysis and Prevention*, Vol. 23, No. 5, 1991, pp. 401-412.

(9) Gaudry, M. and S. Lassarre. *Structural Road Accident Model: The International DRAG Family*. Elsevier Science Ltd., Oxford, 2000.

(10) Bijleveld, F. D. and J. J. F. Commandeur. The Basic Evaluation Model. Presented at the ICTSA Meeting, Arcueil, 2004.

(11) Koopman, S. J. Unobserved Components Time Series Analysis using Stamp7 and X12ARIMA. Presented at Timberlake Consultants Ltd. Course, London, 2005.

(12) Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

(13) Flaig, G. *Unobserved Components Models for Quarterly German GDP*. Publication 681. CESifo, 2002.

(14) Orlandi, F. and K. Pichelmann. *Disentangling Trend and Cycle in the EUR-11 Unemployment Series*. Publication ECFIN/27/2000-EN, No. 140. ECFIN, 2000.

(15) Harvey, A. C. and J. Durbin. The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. *Journal of the Royal Statistical Society A*, Vol. 149, No. 3, 1986, pp. 187-227.

(16) Scuffham, P. A. Economic Factors and Traffic Crashes in New Zealand. *Applied Economics*, Vol. 35, No. 2, 2003, pp. 179-188.

(17) Christens, P. F. *Statistical Modelling of Traffic Safety Development*. Publication IMM-PHD-2003-119. Technical University of Denmark, 2003.

(18) Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering D*, Vol. 82, 1960, pp. 35-45.

(19) Durbin, J. and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, 2001.

(20) Harrison, P. J. and C. F. Stevens. Bayesian Forecasting. *Journal of the Royal Statistical Society B*, Vol. 38, 1976, pp. 205-247.

(21) Sridharan, S., S. Vujic and S. J. Koopman. *Intervention Time Series Analysis of Crime Rates*. Publication TI03-040-4. Tinbergen, 2003.

(22) Koopman, S. J., N. Shephard and J. A. Doornik. Statistical Algorithms for Models in State Space Using SsfPack 2.2. *Econometrics Journal*, Vol. 2, 1999, pp. 113-166.

(23) Doornik, J. A. and M. Ooms. *Introduction to Ox: An Object-Orientated Matrix Programming Language*. Timberlake Consultants Ltd., 2001.

(24) Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kulmala and L. K. Thomsen. Measuring the Contribution of Randomness, Exposure, Weather and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention*, Vol. 27, No. 1, 1995, pp. 1-20.

(25) Van den Bossche, F., G. Wets and T. Brijs. The Role of Exposure in the Analysis of Road Accidents: A Belgian Case-Study. Forthcoming in Transportation Research Record, Journal of the Transportation Research Board, Washington, USA, 2005.

(26) Hakim, S., D. Shefer, A. S. Hakkert and I. Hocherman. A Critical Review of Macro Models for Road Accidents. *Accident Analysis and Prevention*, Vol. 23, No. 5, 1991, pp. 379-400.

(27) Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman. *Applied Linear Statistical Models*. McGraw-Hill, Boston, 1996.

(28) Peltzman, S. The Effects of Automobile Safety Regulation. *Journal of Political Economy*, Vol. 83, No. 4, 1975, pp. 677-725.

(29) Van den Bossche, F., G. Wets and T. Brijs. The Use of Travel Survey Data in Road Safety Analysis. Presented at the European Transport Conference, Strasbourg, France, 2005.

(30) Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. Presented at 2nd International Symposium on Information Theory, Budapest, 1973.

(31) Ljung, G. and G. Box. On the Measure of Lack of Fit in Time Series Models. *Biometrika*, Vol. 65, 1978, pp. 297-303.

(32) Bowman, K. and L. Shenton. Omnibus Test Contours for Departures from Normality Based on Root(b1) and b2. *Biometrika*, Vol. 62, 1975, pp. 243-250.

(33) Van den Bossche, F., G. Wets and T. Brijs. A Regression Model with ARMA Errors to Investigate the Frequency and Severity of Road Traffic Accidents. Presented at 83rd annual meeting of the Transportation Research Board, Washington, USA, 2004.

(34) Zhang, C., J. N. Ivan, W. M. ElDessouki and E. N. Anagnostou. Relative Risk Analysis for Studying the Impact of Adverse Weather Conditions and Congestion on Traffic Accidents. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

(35) Eisenberg, D. The Mixed Effects of Precipitation on Traffic Crashes. *Accident Analysis and Prevention*, Vol. 36, 2004, pp. 637-647.

(36) FEBIAC. Monthly Car Registrations. www.febiac.be. Accessed October 7, 2005.

(37) Box, G. E. and G. C. Tiao. Intervention Analysis with Application to Economic and Environmental Problems. *Journal of American Statistical Association*, Vol. 70, 1975, pp. 70-79.