

Analysis of road risk per age and gender category: a time series approach

Filip A.M. Van den Bossche, Geert Wets¹ and Tom Brijs

Filip A.M. Van den Bossche
Hasselt University
IMOB - Transportation Research Institute
Wetenschapspark 5 bus 6
B-3590 Diepenbeek, BELGIUM
Email: filip.vandenbossche@uhasselt.be
Phone: +32 (0)11 26 91 42
Fax: +32 (0)11 26 91 99

Geert Wets
Hasselt University
IMOB - Transportation Research Institute
Wetenschapspark 5 bus 6
B-3590 Diepenbeek, BELGIUM
Email: geert.wets@uhasselt.be
Phone: +32 (0)11 26 91 58
Fax: +32 (0)11 26 91 99

Tom Brijs
Hasselt University
IMOB - Transportation Research Institute
Wetenschapspark 5 bus 6
B-3590 Diepenbeek, BELGIUM
Email: tom.brijs@uhasselt.be
Phone: +32 (0)11 26 91 55
Fax: +32 (0)11 26 91 99

Date of submission: 2006/08/01

Word count: 5573 + 6*250 = 7073

¹ Corresponding author

ABSTRACT

This paper introduces a road safety analysis for different age and gender categories of road users. In contrast with many previous studies, time series of road crashes per age and gender category will be considered. The objective of the paper is to analyze Belgian data on the yearly number of fatalities per age and gender group, using a decomposition of the number of fatalities in terms of exposure and risk, in a time series perspective. For each category, a state space time series model is developed for the risk, which is defined as the number of fatalities divided by the magnitude of the population. It was found that road risk is changing over the age groups according to a U-shaped curve, and that men generally have a higher risk than women. Further, the risk is decreasing over time, but not at the same rate for all age-gender groups. The highest yearly reduction in risk is found for the oldest and youngest road users. The models are also useful to assess the attainability of formulated road safety targets, which makes them useful policy instruments. Especially for young males, the reduction in risk is not in line with Belgian and Flemish policy expectations.

INTRODUCTION

This paper introduces a road safety analysis for different age and gender categories of road users. In contrast with many previous studies, time series of road crashes per age and gender category will be analyzed. The objective is to enhance the insights in the different characteristics of age and gender groups in terms of their crash involvement, thereby exploring the possibilities of time series analysis by means of state space models.

The approach followed in the present study can be summarized as follows. First, the analysis focuses on road safety for road users of a specific age and gender category. It is often found in literature that road risk varies with age and gender. According to Evans [1], the number of fatalities among drivers in the US shows a peak at the age of 19 years old. The number of fatalities then steadily decreases with age. The completely different behavior in traffic of the age groups will probably be reflected in their crash records. The models developed here are therefore “disaggregated” from a road safety point of view. They describe the road safety situation for various subgroups of road users (per age and gender) and are therefore called “descriptive subset models”.

Second, the number of fatalities will be described in two dimensions, namely the level of exposure and the level of risk. The first dimension, the level of exposure, describes the magnitude of the activity that results in fatalities, and is usually measured in terms of the number of trips, the number of vehicle kilometers, the trip duration or, in the absence of these data, a proxy like the level of the population, as is the case in the study at hand. It accounts for the number of potentially dangerous situations, or the exposure to risk. The second dimension is the probability of a fatality or the risk, given a certain level of exposure. Changes in one of these dimensions will change the safety situation. The two dimensions are naturally related to one another in a multiplicative way:

$$Fatalities = (Exposure) \times \left(\frac{Fatalities}{Exposure} \right) = (Exposure) \times (Risk).$$

In many studies [2, 3], risk is defined as the number of fatalities divided by the level of exposure. When the number of fatalities is studied directly, it is clear that the underlying information on exposure and risk is lost. However, these dimensions can provide useful information on the trends in the number of fatalities. For example, an increase in the number of fatalities can be caused by a rise in the level of exposure, an increase in the risk level or a combination of both. These possibly opposite forces remain hidden if risk is not studied explicitly. Also, if a safety measure leads to a re-distribution of traffic over the different road networks, the safety level may be directly influenced by the changes in exposure. The quantification of these impacts provides useful information for road safety policy makers.

Third, this study takes a time series approach towards exposure and risk, and offers an analysis of road safety for age/gender groups of road users from an evolutionary (or time series) point of view. That is, all data are gathered sequentially and with a regular frequency over time. For example, yearly data on the number of victims for 15-24 years old females can be related to the yearly population level for this group of road users. This also means that the models are aggregated in time. Every observation represents a total number of fatalities for a specific road user group, measured for each year in the time window of the analysis. When analyzing time series data, some interesting questions can be asked that can never be answered with a cross-sectional model. For example, it is possible to calculate a yearly reduction in risk for each group of road users, and to assess the differences between age and gender groups. Also, the risk level and changes in risk over time can be computed.

Fourth, the models are developed for the Belgian road safety situation. For every group of road users with a certain age and gender, both the number of fatalities and the level

of the population will be used. All data are official statistics, gathered by the government and regularly published in official documents. Population data are used instead of vehicle kilometers, because the latter variable is not available for age and gender combinations. However, it is not uncommon to use the population as a base for calculating risk ratios. For example, road safety targets are often formulated in terms of the number of fatalities per 1000000 inhabitants. This study therefore enhances the insights in the Belgian road safety situation, pinpoints the most vulnerable road user groups according to age and gender, and allows comparing the findings with other results.

Based on this view towards the problem, the objective of the paper is to analyze Belgian data on the yearly number of fatalities per age and gender group, using a decomposition of the number of fatalities in terms of exposure and risk in a time series perspective. For each category, a state space time series model is developed for the risk for the road users, which is defined as the number of fatalities divided by the magnitude of the population.

The paper at hand is organized as follows. First, some background on aggregated and disaggregated time series road safety models is given. Then, the data sources and the methodology of state space models are introduced. Next, the main results are presented. The paper ends with some general conclusions on the use of these models, and some topics for further research.

BACKGROUND

In this paper, road fatalities and population per age and gender group are analyzed in this paper by means of time series analysis. Investigating time series of road safety data is, for various reasons, a rewarding effort. Governments are more than ever concerned with the negative consequences of road traffic. The loss of human lives and economic capital due to road accidents forced governments to plan and execute road safety actions [4]. Until the early seventies, the planned actions were of a reactive nature. The main objective was to stop or slow down the negative developments in road safety. Afterwards, the focus changed to more strategic actions. It was commonly accepted that strategic plans for future safety improvements had to be developed, based on the knowledge of the past.

This change in policy is mainly characterized by the introduction of strategic plans, together with explicit road safety targets. It is clear that, by putting quantitative targets, the noncommittal attitude towards the road safety problem is no longer accepted and underlines the need for ambitious, yet realistic targets. In order to set these targets and to specify or adapt safety plans accordingly, it is necessary to measure the developments in road safety and to understand the underlying processes of exposure and risk. This is the main motivation for the use of time series analysis techniques in road safety research. A sound analysis of the evolution in time is needed to set future road safety targets and to assess the efforts made to achieve them.

A strategic road safety device allows describing and monitoring the number of fatalities in road crashes, by looking for example at trends over the years. Many models have been developed to explain the long-term evolution in the number of fatalities [4]. Aggregated models, usually developed on a countrywide level for the total number of crashes or fatalities, allow gaining a better insight in the evolution of fatalities over time. The annual number of fatalities is an important indicator of the level of road safety in a country. Policy makers use these statistics to indicate past trends and to create rough indicators of the future evolution. In many countries, the government formulated clear long-term quantitative objectives in terms of the number of fatalities. In Belgium and in Flanders, as well as on the European level, the goal is to half the total number of fatalities by 2010. The Belgian States General for Road Safety [5] recommended the objective of halving the number of fatalities by 2010, compared

to the average number of fatalities for the years 1998-2000. This boils down to a maximum of 750 fatalities in 2010. Also, the maximum value of the number of fatalities per 100000 inhabitants in 2010 is set at 7. In Flanders, the government expects making up arrears compared to the best performing countries in Europe by 50%, taking into account the own ambitions of these leading countries in terms of road safety. Broadly speaking, this implies a reduction in the number of fatalities by 50% compared to 1999, or a maximum of 375 fatalities in 2010 [6]. Another Flemish objective for the year 2010 is a 60% reduction in the number of young persons (age below 26) killed per 1000000 youngsters. In [7], an overview is given of the quantitative targets formulated in some OECD countries. These targets are in terms of annual fatalities (either in absolute values or related to the level of the population or the number of vehicle kilometres), and should therefore be supported by models at this level of aggregation.

When road safety has to be analyzed for specific parts of the traffic system in a time series context, disaggregated models can be used. In the COST-329 project of the European Commission [4], disaggregated models are defined as “models in which the response (dependent) variable comprises a sub-group of the total, aggregated number of accidents or their consequences”. Sometimes, the term “disaggregated” is used for models in which data pertaining to individual units or decision makers (cars, travelers, households, firms, etc.) are analyzed. The models presented here could therefore rather be named “accident or casualty subset models”. From a time perspective, however, the models are still aggregated, in the sense that they consider total numbers of accidents or victims, aggregated per year, but now for a subset of the whole road safety system. Typical subset examples are types of roads (highways, rural roads and urban roads), types of road users (according to their age and/or gender) or vehicle types involved in a crash. Depending on the data available, these subgroups can be crossed-over, by looking for example at the fatalities of female road users on highways or children injured in the urban area. It is clear that with more restricted groups of road users, one runs faster into data problems.

Subset models are useful in addition to the aggregated models that consider the road safety situation as a whole. Road safety models are meant to support decision- and policy makers in their analysis of road safety developments, especially when setting road safety targets and developing road safety plans. While an aggregate model will typically be used for the description and forecasting of general trends in road safety on a high level, they are less suited for analyzing parts of the transport system or subgroups of road users. For example, changes in the proportion of young and old road users may affect the evolution in road safety. It is not possible to derive the effects of these changes from an aggregated model. Therefore, it is necessary to analyze the trends in road safety at a lower level of aggregation.

In the COST 329 report [4], an overview of several examples of disaggregated road safety modeling are given. Broughton [8] computed trends for casualty rates per type of road user. On the basis of population forecasts, these trends are further disaggregated according to age and gender. In Bijleveld & Oppe [9], the number of fatalities and rates for combinations of transport mode and age, and separately for types of road, are analyzed. Later, multivariate nonlinear time series model for the analysis of traffic volumes and road casualties inside and outside urban areas were developed by Bijleveld et al. [10]. Greibe [11] gives a description of the number of killed and seriously injured cyclists on urban roads using a log-linear model including variables on bicycle and motor vehicle traffic and on general safety improvements. Pedersen [12] uses log-linear splines to detect abrupt changes in the development of casualties. According to the author, this approach can be applied at every level of aggregation. Also some of the DRAG models [13] can be seen as disaggregated models, in the sense that they consider aspects of specific groups of road users, like for example in [14].

For Belgium, disaggregated models were developed by age and gender, type of road, type of road user and crash type [15].

Although it is clear that studies at a lower aggregation level should enhance the insights in the road safety trends for parts of the road system, the number of subset time series models in road safety analysis is quite small. This may be explained by the high data needs that are typical for these models. The lower the level of aggregation, the more difficult it becomes to obtain the data needed for the analysis. While aggregated statistics on road safety and exposure are generally available, they are mostly not easily found for subgroups of road users. Also, due to the smaller counts on a lower aggregation level, the data might be less accurate and more advanced statistical techniques are needed to adequately model the problem. The higher data requirements, especially in a time series context, can probably explain the relatively low number of disaggregated time series studies. On the other hand, it is not illogical to consider the risk according to the age and gender of the road users. It is known that young persons are often involved in road crashes. Looking at the official statistics [16], road crashes are clearly one of the most important death causes for young persons. In 2001, approximately 36% of the Belgian fatal road victims was between 10 and 29 years old.

DATA SOURCES

For Belgium, data are available on the number of fatalities per age and gender category, as well as on the magnitude of the population for these groups. The population is a key indicator of the Belgian economy and is maintained and published by the Belgian National Bank. As mentioned before, the population data will be used to calculate risk measures for each age and gender group. Although it is not a perfect measure of exposure (nothing is said about distances traveled, for example), the population data provide an upper bound of the number of road users. Given the assumption that not the whole population travels (especially for older people), the risk of being killed will be underestimated. However, current changes in the structure of the population might be reflected in the data.

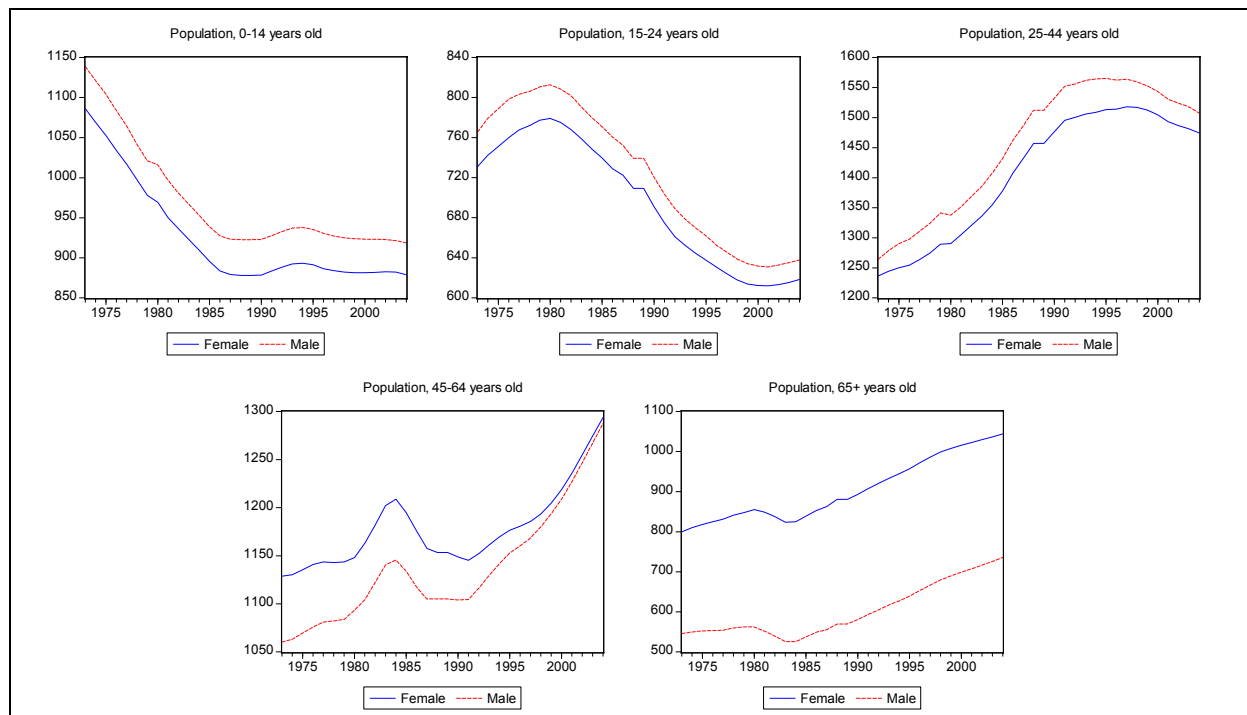


Figure 1: Population data for various age/gender groups

The population levels per gender for five age categories (0-14, 15-24, 25-44, 45-64 and 65 and older) are given in Figure 1 for the years 1973-2004. It can be seen that the youngest group of road users has been declining for about 25 years, and seems to be stabilizing in the late nineties. The group of older people is growing. It is also interesting to note that females are the smallest group among the younger people, and the largest group among the older. It is clear that these changes in the population might also be reflected in the level of exposure to road crashes.

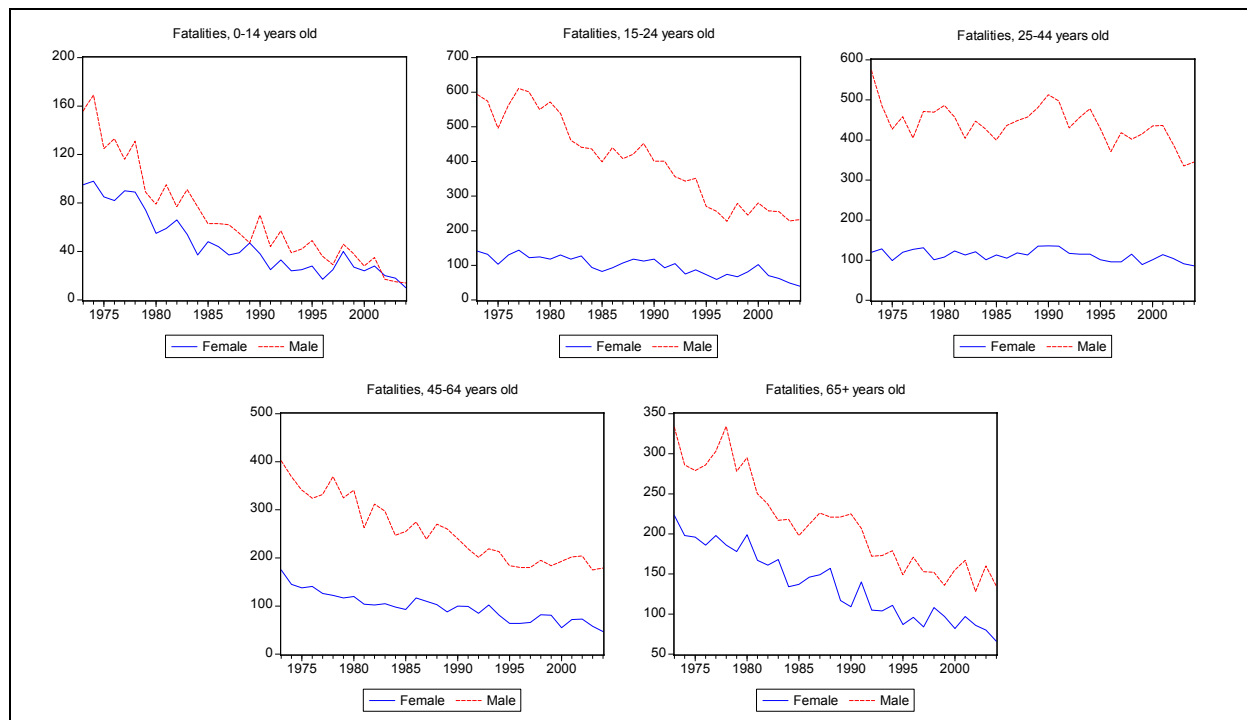


Figure 2: Annual number of fatalities by age and gender

The road safety data are obtained from the National Institute of Statistics. In Figure 2, the number of fatalities is shown for each age-gender category. Data are available for the period 1973-2004 for the same five age groups. The choice for the categories is dictated by the availability of the data.

The evolution in time of the fatalities per age and gender category shows some very specific properties. First, the number of fatalities is higher for males in every age category. This is true for the complete time window of the study. However, the differences get smaller for all categories, except for the 25-44 years old victims. Second, the number of fatalities is decreasing over time, except for the 25-44 years old victims. For the latter group, the level of fatalities hardly changed.

METHODOLOGY

The data on population and fatalities will be analyzed by means of state space models. Like the classical ARIMA models, developed by Box and Jenkins [17], state space models can be considered as dedicated time series models. That is, they are specifically developed to take time dependencies into account. State space models are typically handled with the Kalman filter, a method of signal processing which provides optimal estimates of the current state of a dynamic system. An observation y_t of a series at time t is written in state space notation as [18]:

$$y_t = h_t^T \theta_t + n_t.$$

The state vector θ_t is usually unobservable and contains the state variables as components. The state variables are typically model parameters, like regression coefficients, or parameters describing the state of a system (such as the level). The column vector h_t is known, and n_t is the observation error. It is further assumed that the state θ_t depends on the previous state θ_{t-1} , and that the changes of θ_t through time follow the equation:

$$\theta_t = G_t \theta_{t-1} + w_t.$$

Here, G_t is assumed to be known, and w_t denotes a vector of white noise deviations. Both equations together form the state space model. The first equation is called the observation equation, while the second is the state equation. The errors are generally assumed to be serially uncorrelated and normally distributed.

After formulating the model in terms of its components, the main objective usually consists of estimating the signal, represented by θ_t . The Kalman filter can be used to estimate this unobserved vector. The ‘‘Kalman recursion equations’’ enable the calculation of the one-step forecast errors and the likelihood [19]. This is usually done in two stages, as described for example in [18]. In the prediction stage, θ_t is forecasted from the data up to time period $(t-1)$. When subsequently the new observation at time t has been observed, the estimator for θ_t can be modified to take account of this extra information. The prediction error of the forecast of y_t is used to update the estimate of θ_t . This is the updating stage of the Kalman filter. The advantage of the recursive character of the Kalman filter is that every new estimate is based on the previous estimate and the latest observation, while at the same time the whole past of the series is taken into account.

The theory of state space models and their applications are extensively described in basic references such as [20] and [21]. The models that will be considered in this text for each age and gender combination can, in their most general form, be written as follows:

$$\begin{aligned} y_t &= \mu_t + \sum_{j=1}^J \beta_{j,t} x_{j,t} + \varepsilon_t & \varepsilon_t &\sim NID(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + v_t + \xi_t & \xi_t &\sim NID(0, \sigma_\xi^2) \\ v_{t+1} &= v_t + \zeta_t & \zeta_t &\sim NID(0, \sigma_\zeta^2) \\ \beta_{j,t+1} &= \beta_{j,t} & j &= 1, \dots, J. \end{aligned}$$

In this expression, the first equation is the observation equations, while the second and the third are state equations. Further, μ_t denotes the level and v_t is the slope. Just like in classical regression, the slope determines the rate of change in the state, but here it is allowed to change over time. The combination of the level and the slope component determines the trend in the model. The effects of the intervention variables $x_{j,t}$ are measured by the parameters $\beta_{j,t}$. Possible interventions considered here are pulse interventions (to take outliers into account) and level shifts. In the equation above, it is assumed that the regression parameters for the interventions do not change over time, although the framework is easily extended to do so. The error terms in the model are all assumed to be normally and independently distributed with zero mean and a specific standard deviation, denoted σ_ε for the observation equation and σ_ξ and σ_ζ respectively for the level and the slope. These standard deviations can be used to assess the assumption that a component is indeed stochastic. If the estimation procedure shows that the standard deviation of a specific component is not significantly different from zero, it can be decided to model it as a fixed component. In such a way, it is clear that deterministic models, like the classical regression models, are special cases of the stochastic state space model.

RESULTS

The model presented in the methodology section of the paper is used to analyze the log of the risk (y_t) for each age and gender combination (5 age categories and 2 gender categories), defined as the number of fatalities divided by the magnitude of the population. The index t specifies the year of the observation ($t = 1973, \dots, 2004$). The observation equation and the equations for the level and the slope each have their own irregular term, ε_t , ζ_t and ξ_t . These terms are assumed to be uncorrelated and normally distributed. To test for uncorrelated residuals, the Box-Ljung Q-statistic is used at different lags. This statistic is given by [19]:

$$Q_K = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k}.$$

Here, r_k is the lag- k autocorrelation coefficient, n is the number of observations and K is the maximum lag being considered. Under the null hypothesis of no serial correlation, this test statistic is asymptotically distributed as $\chi^2(K-M)$, where M equals the number of parameters estimated in the model. Large autocorrelation coefficients lead to a high Q-statistic. A high value therefore indicates significant autocorrelation and thus rejection of the null hypothesis. To test for normality, the Bowman-Shenton test is used. This test takes into account the skewness and the kurtosis of the distribution of the residuals and is tested against a $\chi^2(2)$ distribution [22].

Table 1: Results for the age-gender models

Age	Gender	Param.	Coeff.	R.m.s.e.	t-value	Prob.
0 – 14	Male	μ	-4.0601	0.1082	-37.523	0.0000
		ν	-0.0685	0.0168	-4.064	0.0003
		β_{1990}	0.3033	0.1272	2.3842	0.0236
		β_{1998}	0.2991	0.1310	2.2822	0.0297
		β_{2002}	-0.5372	0.0893	-6.0141	0.0000
	Female	μ	-3.9935	0.0683	-58.444	0.0000
		ν	-0.0534	0.0037	-14.295	0.0000
		β_{1998}	0.5795	0.1982	2.9247	0.0065
15 – 24	Male	μ	-1.0137	0.0438	-23.157	0.0000
		ν	-0.0239	0.0112	-2.1329	0.0412
	Female	μ	-2.7015	0.0673	-40.16	0.0000
		ν	-0.3363	0.0267	-1.2613	0.2169
25 – 44	Male	μ	-1.4803	0.0368	-40.273	0.0000
		ν	-0.0210	0.0125	-1.6714	0.1050
	Female	μ	-2.7639	0.0437	-63.237	0.0000
		ν	-0.0141	0.0038	-3.6941	0.0009
45 – 64	Male	μ	-1.9662	0.0404	-48.678	0.0000
		ν	-0.0303	0.0050	-6.0751	0.0000
	Female	μ	-3.1761	0.0714	-44.47	0.0000
		ν	-0.0397	0.0123	-3.2337	0.0030
65+	Male	μ	-1.6629	0.0492	-33.773	0.0000
		ν	-0.0363	0.0082	-4.4282	0.0001
	Female	μ	-2.6230	0.0418	-62.703	0.0000
		ν	-0.0429	0.0035	-12.319	0.0000

Whenever necessary, intervention variables were introduced into the models to render the residuals uncorrelated and normally distributed. This was the case in the risk models for

the first age category (pulse intervention in 1998 for females; pulse interventions in 1990 and 1998, level shift in 2002 for males). For the final models, the Q-statistics did not indicate any problem of autocorrelation and the null hypothesis of normally distributed residuals in the observation equation could not be rejected.

The variances of the residuals for the level and the slope equations were used to assess whether these components should be stochastic or deterministic. If the variance of one of these residuals is estimated to be equal to zero, the component is assumed to be deterministic. For the first age category, the models have a deterministic level and slope, which corresponds to the classical regression model. For the other models, a stochastic level with a deterministic slope was retained. Estimation is done in STAMP 6.21 [22]. The model is estimated for each age-gender combination, leading to 10 separate equations. The results are shown in Table 1.

The parameter estimates for μ and ν are for the level and the slope respectively, measured at the end of the sample period (2004). As the slope is fixed in all models (that is $\sigma_\zeta = 0$), it is also the slope for the entire analysis period. This is not the case for the trend. The β values are the parameter estimates for the included intervention variables. These interventions were only needed for the first age category. There is no specific reason for these interventions, apart from a drastic restructuring of the data gathering process in 2002.

It is further seen from the table that most of the parameter estimates are significantly different from zero, except for the slope in two models (15-24 Female and 25-44 Male). As the slope is fixed and not statistically different from zero, the risk for these groups is almost constant over time. The slopes in Table 1 can be used to calculate a yearly reduction in risk over time for each age and gender category. This is illustrated in Figure 3. This information learns that, for example in the group of male road users between 15 and 24 years old, the average decrease in risk equals $1 - \exp(-0.0239)$ or 2.39%. A high value indicates a higher reduction in risk over time. A U-shaped relation is found for males and females, indicating a higher risk reduction per year for younger and older people. Clearly, the most “active” age groups have the lowest yearly reduction in risk compared to the other groups.

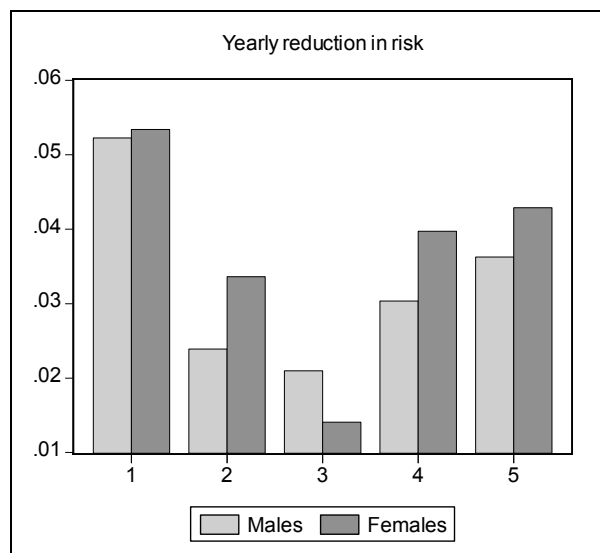


Figure 3: Yearly reduction in risk for males and females

Another interesting output from these models is the risk pattern over the different age groups. This is illustrated in Figure 4. The graphs show the estimated risk for male and female road users for each age category for the years 1980, 1990 and 2000, calculated from the 10 models. The risk curves show a U-shaped pattern, starting from the second age category. The same shape is found by Evans [1] for population fatality rates of drivers

between 20 and 80 years old, although his analysis was not over time. As in our analysis the age of 20 belongs to the second category, the U-shaped relation found here is comparable to that found by Evans. It is also seen that the risk is generally lower for females than for males, although the difference is quite small in the first age category. Further, the three curves in each graph indicate the reduction in risk over time. Clearly, the decrease in risk is not equally large for every age-gender combination, which underlines the importance of targeted road safety programs and campaigns.

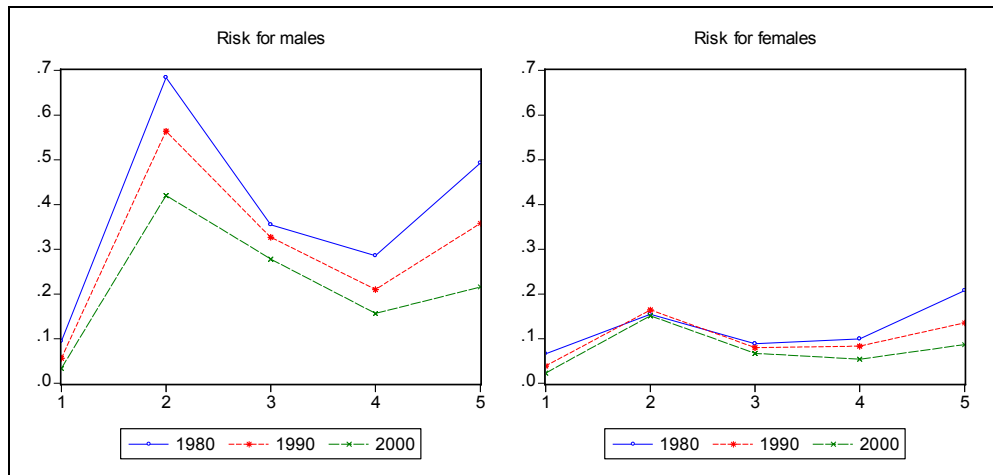


Figure 4: Risk comparison over time for males and females

As mentioned before, road safety objectives are often expressed in terms of a reduction in the number of fatalities per inhabitant. The models presented here can be used to assess the reduction that is expected to be achieved by 2010 for the various age and gender groups. To do so, the risk level is predicted 6 years ahead, from 2005 up to 2010, such that the predicted value for 2010 can be compared with, for example, the observed risk level in 2000.

Table 2: Road safety objectives assessment

Age	Gender	Observed risk 2000	Estimated risk 2010	Estimated reduction
0 – 14	Male	0.0303	0.0115	-62.036%
	Female	0.0272	0.0134	-50.845%
15 – 24	Male	0.4432	0.3144	-29.074%
	Female	0.1666	0.0548	-67.079%
25 – 44	Male	0.2819	0.2007	-28.815%
	Female	0.0671	0.0579	-13.684%
45 – 64	Male	0.1597	0.1167	-26.921%
	Female	0.0451	0.0329	-27.107%
65+	Male	0.2217	0.1525	-31.209%
	Female	0.0807	0.0561	-30.468%

This exercise is done in Table 2. For every age and gender group, the table shows the expected reduction in 2010 that will be obtained when the current road safety policies are continued. First, the table shows large differences between males and females, especially for the second and third age category. Although the difference is clearly present, objectives are rarely formulated towards a specific gender. Second, huge differences are to be noted

between the age categories. Youngsters groups show the highest reductions in risk. In the first two age categories (0-24), the risk is reduced by more than 50% for all groups, except for the males in the second group. This result indicates that young males form a high risk group, especially at the age where they start driving cars, have a more independent social life and start working. In the targeted group of youngsters, males of 15-24 years old deserve special attention.

CONCLUSIONS AND FURTHER RESEARCH

In this paper, the risk for different age and gender categories of road users is studied. Road risk, expressed as the number of fatalities per population level, is analyzed by means of state space models. It was found that road risk is changing over the age groups according to a U-shaped curve, and that men generally have higher risk than women. Further, the risk is decreasing over time, but not at the same rate for all age-gender groups. The highest yearly reduction in risk is found for the oldest and youngest road users. The models are also useful to assess the attainability of formulated road safety targets. Especially for young males, the reduction in risk is not in line with policy expectations. Although the models presented here are quite simple, they offer the possibility of describing the differences in risk between various age-gender categories, together with their trend over time. The results show that road user groups have very specific characteristics, which are worth investigating separately. Some road safety measures are only oriented towards one or more subsections of the transport system, and testing these measures on a highly aggregated level might obscure the nature of the effects.

The field of subset time series modeling in road safety research seems to be scarcely out of the egg. However, the merits of this approach are clear. A disaggregated analysis is complementary to the aggregated models in the sense that the trends for specific parts of the transport system can be analyzed separately, and the specific countermeasures, introduced to enhance road safety for a specific group of road users, can be benchmarked.

The subset time series modeling approach is mainly restricted in its applicability by the lack of appropriate data, at least for the Belgian situation. First, the time series models can only be developed for the yearly total number of fatalities per age and gender category. That is, no further details concerning the type of road user (car, bicycle, motorcycle, etc.) can be analyzed. Given the importance of vulnerable road users in traffic, it is desirable to investigate the differences in road safety between these categories. Second, for age and gender categories, no time series of exposure, in terms of vehicle kilometers, are available. In the time series models, population data are used instead. Using population data as a measure of exposure will perhaps not give the same results as when the number of kilometers driven is used. Changes in population will, for some age and gender categories (especially for older and younger persons), diverge from the number of kilometres driven by these groups of road users. Typically, one can expect that older people drive less than younger ones, although their share in the total population is rising. However, Evans [1] found comparable U-shaped curves for driver fatalities per billion kilometer. This confirms the fact that it is useful to include population data to correct for the differences between the age groups over time. The fact that the models can be used to assess road safety policy targets makes them very attractive policy instruments.

REFERENCES

- (1) Evans, L., *Traffic safety*. Science Serving Society. 2004, Michigan: Bloomfield Hills. 445.
- (2) Lassarre, S., *Analysis of progress in road safety in ten European countries*. Accident Analysis and Prevention, 2001. **33**: p. 743-751.
- (3) Oppe, S., *The development of traffic and traffic safety in six developed countries*. Accident Analysis and Prevention, 1991. **23**(5): p. 401-412.
- (4) COST 329, *Models for traffic and safety development and interventions*. 2004, Directorate General for Transport, European Commission: Brussels, Belgium. p. 235.
- (5) Staten-Generaal voor de Verkeersveiligheid. 2001, <http://www.wegcode.be>.
- (6) Ministerie van de Vlaamse Gemeenschap, *Ontwerp Mobiliteitsplan Vlaanderen*. 2001, Ministerie van de Vlaamse Gemeenschap.
- (7) Van Schagen, I., *De verkeersonveiligheid in Nederland tot en met 1999: analyse van aard, omvang en ontwikkelingen*. 2000, Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV. p. 92.
- (8) Broughton, J., *Disaggregate casualty forecasting*. 1994, Road Safety Division, DETR, reproduced in COST-329 reader: Great Britain.
- (9) Bijleveld, F.D. and S. Oppe, *Time-ordered analysis of disaggregate road safety data*, in *COST-329 reader*. 1999, SWOV: Leidschendam, The Netherlands.
- (10) Bijleveld, F.D., et al., *Multivariate nonlinear time series modelling of exposure and risk in road safety research*. 2005, Vrije Universiteit: Amsterdam.
- (11) Greibe, P., *Cyclist safety in urban areas*. 1999, Danish Road Directorate, Ministry of Transport: Denmark.
- (12) Pedersen, C.B., *Log linear splines*, in *COST-329 reader*. 1998, European Commission, Directorate General for Transport: Brussels, Belgium.
- (13) Gaudry, M. and S. Lassarre, *Structural road accident models: the international DRAG family*. 2000, Oxford, UK: Elsevier Science Ltd.
- (14) Fridstrøm, L., *The TRULS-1 model for Norway*, in *Structural Road Accident Models: the International DRAG Family*, M. Gaudry and S. Lassarre, Editors. 2000, Elsevier Science: Oxford. p. 97-126.
- (15) Van den Bossche, F., *Road Safety, Risk and Exposure in Belgium: an econometric approach*, in *Applied Economics*. 2006, Hasselt University: Diepenbeek, Belgium. p. 233.
- (16) D'Hondt, H., *Gezondheid - Doodsoorzaken in 1997 volgens de aangiften van overlijden*. 2002, Nationaal Instituut voor de Statistiek: Brussels, Belgium.

- (17) Box, G.E.P. and G.M. Jenkins, *Time series analysis: forecasting and control*. 1976, San Francisco, USA: Holden-Day. 575.
- (18) Chatfield, C., *The analysis of time series - An introduction*. 6 ed. Texts in statistical science. 2004, Boca Raton, Florida: Chapman & Hall/CRC.
- (19) Makridakis, S., S.C. Wheelwright, and R. Hyndman, *Forecasting: methods and applications*. 3 ed. 1998: John Wiley & Sons, Inc. 656.
- (20) Harvey, A.C., *Forecasting, structural time series models and the Kalman filter*. 1989, Cambridge, United Kingdom: Cambridge University Press. 555.
- (21) Durbin, J. and S.J. Koopman, *Time series analysis by state space methods*. Oxford statistical science series, ed. A.C. Atkinson, et al. Vol. 24. 2001, Oxford, UK: Oxford University Press. 253.
- (22) Koopman, S.J., et al., *Stamp: structural time series analyser, modeller and predictor*. 2000, Timberlake Consultants Ltd.: London, UK.