

Deliverable D7.10:

Time series applications on road safety developments in Europe

Please refer to this report as follows:

Stipdonk, H.L. (Ed.) (2008) *Time series applications on road safety developments in Europe*. Deliverable D7.10 of the EU FP6 project SafetyNet.

Contract No: TREN-04-FP6TR-S12.395465/506723

Acronym: SafetyNet

Title: Building the European Road Safety Observatory

Integrated Project, Thematic Priority 6.2 "Sustainable Surface Transport"

Project Co-ordinator:

Professor Pete Thomas

Vehicle Safety Research Centre

Ergonomics and Safety Research Institute

Loughborough University

Holywell Building

Holywell Way

Loughborough

LE11 3UZ

Organisation name of lead contractor for this deliverable:

SWOV Institute for Road Safety Research

Editor:

H.L. Stipdonk (SWOV)

Report Author(s): J.J.F. Commandeur, F.D. Bijleveld, Y. van Norden (SWOV), R. Bergel, M. Debarh (INRETS), G. Yannis, C. Antoniou, D. Katsochis, E. Papadimitriou (NTUA)

Due Date of Deliverable: 31/10/2008

Submission Date: 31/10/2008

Project Start Date: 1st May 2004

Duration: 4.5 years

Dissemination Level	
PU	Public



Project co-financed by the European Commission, Directorate-General Transport and Energy

Executive summary

This report shows how traffic safety data can be analysed to understand the development of traffic safety over time.

The report shows a number of principles of time series analysis:

- The current state of the art of structural time series analysis is described. We illustrate the importance of changes in distance travelled, and how this affects the number of accidents or the number of fatalities. Preferably, current methods simultaneously analyse the development of distance travelled, and of the number of accidents or fatalities, so as to allow for robust forecasts.
- Further understanding of the development of traffic safety asks for disaggregation of the safety data into different subgroups. Such disaggregation reveals different trends for different groups. An analysis of the development over time of mortality (fatalities per inhabitant) is given for different European countries. This is compared to motorization rate. The analysis shows how motorization rate and mortality relate, and what different patterns evolve when countries are compared.
- The development of traffic safety is shown to depend strongly on the composition of the population of a country. It is well known that young, unexperienced drivers are at higher risk, especially male drivers. Changes in the demographics are related to safety, and forecasts are sensitive to this composition. We show how to use population data to analyse and forecast the number of fatalities in a country.
- Eventually, models are expected to incorporate other influencing factors or safety performance indicators, such as the quality of roads, fraction of safety belts used etc. At this stage, operational models of that type are a bridge too far, mostly, because most influencing factors are very specific. As an example of a factor that affects all accidents, we chose weather parameters to incorporate in our models. We use two different techniques to analyse the effect of precipitation, frost and temperature on road safety.

Part one of the report gives a bird's-eye view of the time series analyses that were carried out in the final year of SafetyNet. We did our best to present the main results in such a way that non-technical readers can easily follow the lines of thought. The main lines of reasoning are given and explained, and the highlights of the results are presented.

Part two consists of four separate contributions, each giving a more detailed description of the objectives, methods and results of each contribution. This part is more technical, but still sufficiently accessible to non-technical readers to get the grip of the techniques described.

Part three is a list of appendices, where technical and mathematical details can be found, or large sets of graphical results are presented. This part is especially meant for the very interested or adepted readers.

Table of Contents

CHAPTER 1 - BIRD'S EYE VIEW	11
1.1 Introduction.....	11
1.2 The three levels of risk.....	14
1.3 Modelling road fatality trends in the European countries	15
1.4 Describing trends by disaggregation by age and gender.....	17
1.5 Modelling meteorological effects on road safety	21
1.5.1 The Athens region.....	21
1.5.2 Modelling the weather effect in different regions.....	24
CHAPTER 2 - MODELLING THE THREE LEVELS OF THE ROAD RISK PROCESS AT AN AGGREGATE LEVEL.....	29
2.1 Introduction.....	29
2.2 Method.....	30
2.3 Data.....	31
2.4 Results	32
2.5 Discussion and perspectives	37
2.6 References	40
CHAPTER 3 - MODELLING ROAD FATALITY TRENDS IN THE EUROPEAN COUNTRIES	41
3.1 Introduction.....	41
3.2 Data and application	43
3.3 Methodology	46
3.4 Estimated models.....	47
3.5 Conclusion	53
3.6 References	56
CHAPTER 4 - DISAGGREGATION BY AGE AND GENDER.....	59
4.1 Introduction.....	59

4.2 Decomposition of mobility into population size and mobility per capita	60
4.3 Forecasting victim counts per age class and gender	66
4.3.1 Technical details of the model	67
4.3.2 Highlights of the resulting forecasts	68
4.4 Conclusion	68
4.5 References	69
CHAPTER 5 - MODELLING METEOROLOGICAL EFFECTS ON ROAD SAFETY.....	71
5.1 General introduction	71
5.2 Impact of meteorological variables on the number of injury accidents and fatalities in the Athens region.....	72
5.2.1 Introduction.....	72
5.2.2 Data used.....	72
5.2.3 Daily analysis without exposure data.....	75
5.2.4 Results.....	76
5.2.5 Conclusion	78
5.3 Modelling the weather effects on the numbers of injury accidents at an aggregate level in different regions	79
5.3.1 Introduction.....	79
5.3.2 Methodology and data.....	79
5.3.3 Results.....	80
5.3.4 Conclusion	82
5.4 General discussion	83
5.5 References	83
APPENDICES.....	89
Appendix A2.1.....	89
Appendix A2.2.....	91
Appendix A3.1.....	93
Appendix A4.1.....	95
Appendix A4.2.....	133
Appendix A4.3.....	134
Appendix A4.4.....	139
Appendix A5.1.....	144

Appendix A5.2..... 145

Part I

Chapter 1 - Bird's eye view

Jacques Commandeur, Yvette van Norden (SWOV)

1.1 Introduction

With time series data of road safety and their analysis, interesting questions can be asked that can never be answered with a cross-sectional model. Cross-sectional data and their analysis typically provide a 'frozen' snap-shot of the road safety situation at a fixed point in time. On the other hand, only with time series data is it possible to study *changes* in exposure, risk, and road accidents and their outcomes over time, and to provide estimates of their future developments, thus helping policy makers in developing realistic quantitative safety targets.

Examples of analyses of road safety data using time series analyses are presented in this Deliverable, to illustrate how these models can be used to explain quantitatively the developments of road safety. The variables we studied (driver age, gender, traffic mode and weather) were chosen because of feasibility of the analyses, considering the complexity of the models and the availability and quality of data. In the future, hopefully many relevant other variables can be studied as well after improvement of the data availability and quality where necessary, and as a result enable further comparison between countries.

We start by sketching a very global picture of the development of road safety in the European Union as a whole, as illustrated in Figures 1.1 to 1.3. In these figures we have plotted the total number of traffic fatalities, the total size of the vehicle fleet, and their ratio's as observed in 16 of the 27 current member states of the European Union in the years 1971-2004. Information on the countries Bulgaria, Czech Republic, Cyprus, Estonia, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia, and Romania is missing from these figures due to the unavailability of (some of) the required data.

In Figure 1.1 we see that overall the number of road traffic fatalities in the European Union has been decreasing during the last decades, except for a few temporary increases in 1972 and at the end of eighties. At the same time, the total number of vehicles has continuously and almost linearly increased as can be deduced from Figure 1.2. The annual ratios of the fatalities in Figure 1.1 and the number of vehicles in Figure 1.2 are given in Figure 1.3. Figure 1.3 fortunately indicates that overall the risk of being killed in road accidents in the European Union has continuously decreased during the last decades.

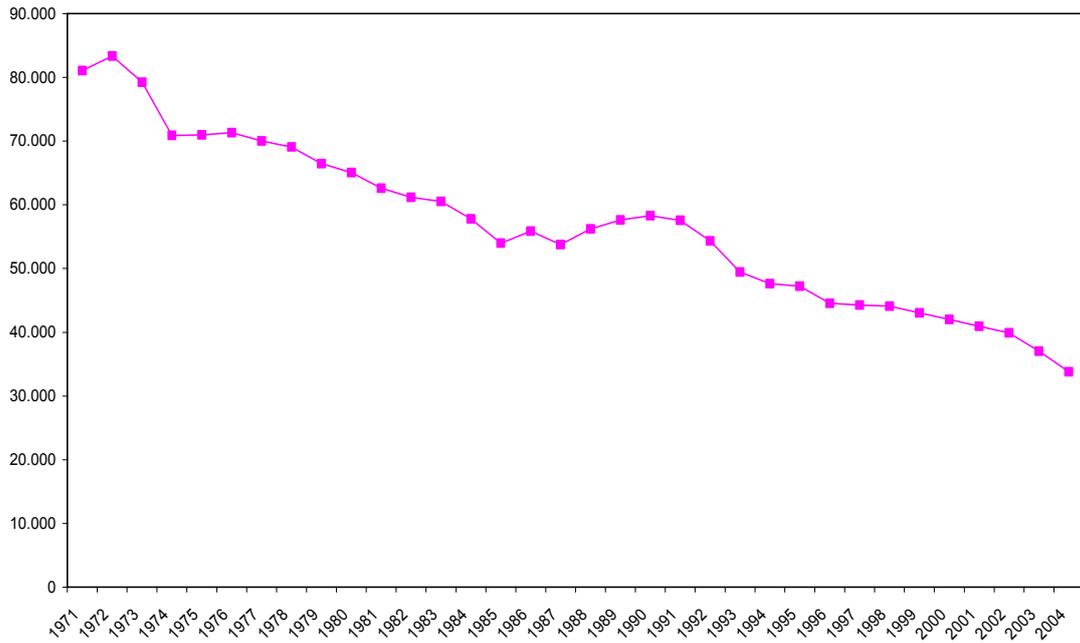


Figure 1.1: Total number of fatalities in 16 EU countries.

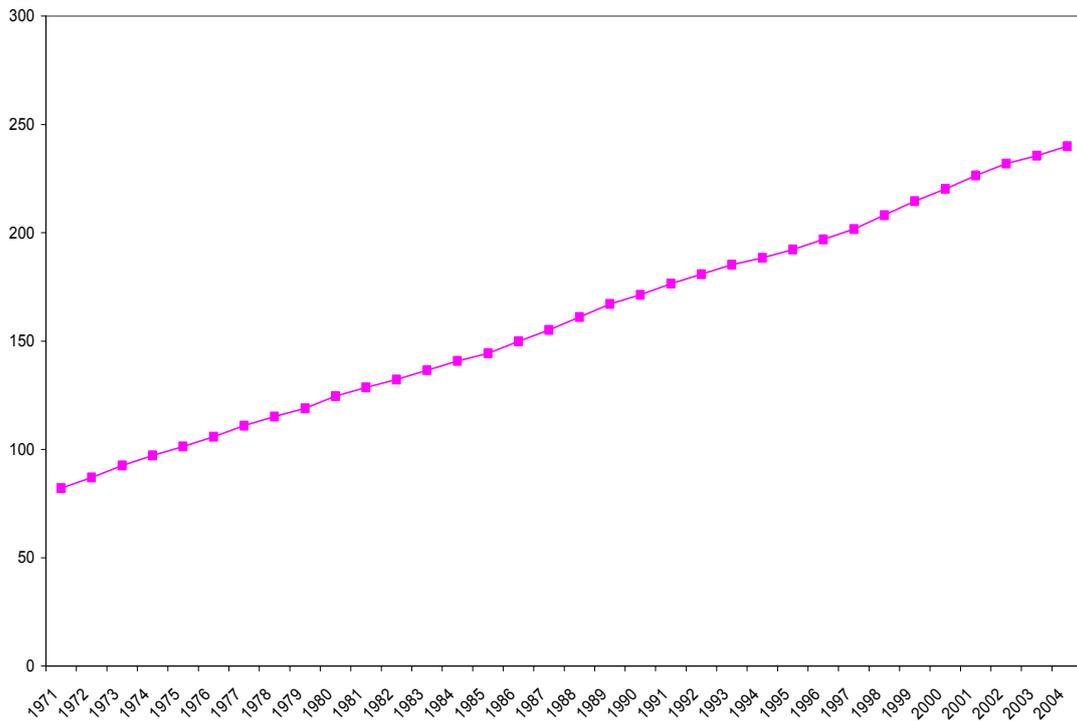


Figure 1.2: Total vehicle fleet (in millions) in 16 EU countries.

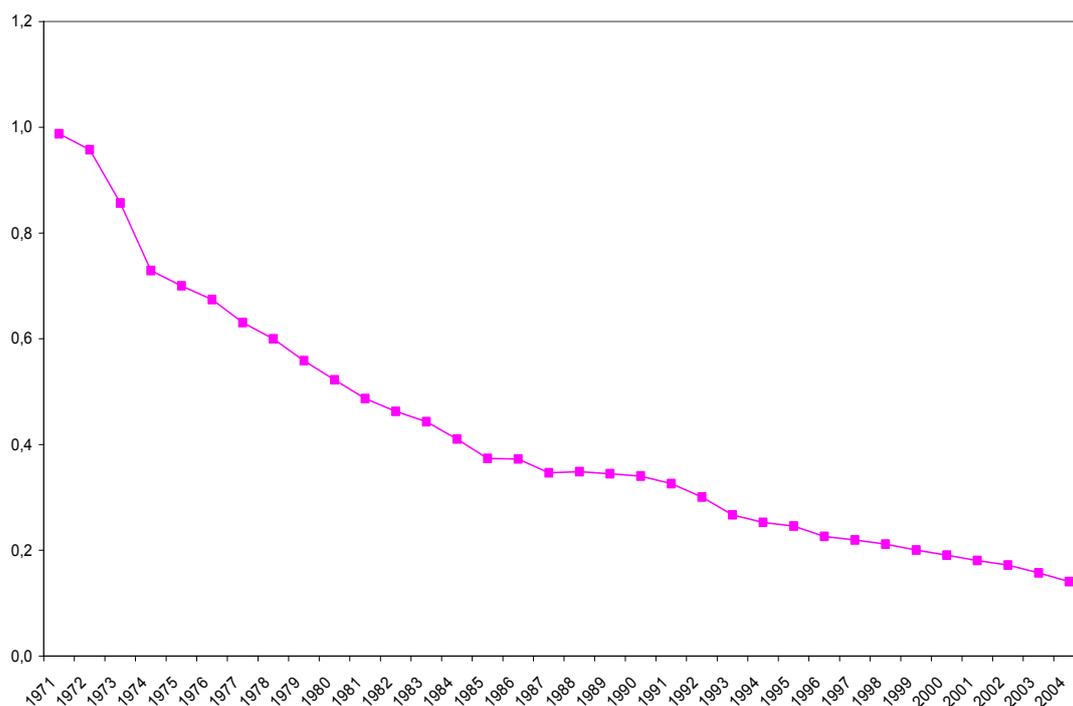


Figure 1.3: Total fatality rate (killed per thousand vehicles) in 16 EU countries.

However, informative as these three figures may be in providing a first impression of the main changes that occurred during the last decades in road safety at the global level of the European Union, they are not very helpful when it comes to understanding *why* and *how* these changes in road safety came about.

There are two important ways in which we can get insight into the factors explaining the changes in these global developments. The first is to include *generic* explanatory variables in the modelling of these developments. In other words, variables which may explain changes in the aggregate number of accidents or fatalities are incorporated in the models. Examples of these variables are changes in road infrastructure in general, the quality of all vehicles, law regulations, the enforcement of these regulations, weather conditions, et cetera. Then we investigate whether the changes in these explanatory variables contribute significantly to the observed changes in exposure, risk and/or road accidents and their outcomes. The second way to obtain a better understanding of the observed changes in these global developments is to disaggregate (or stratify), that is to investigate the developments in subgroups of the total traffic process. This allows for *specific* variables describing changes in subgroups. Moreover, it is only by studying disaggregated developments that we can find out which subgroups performed better or worse than others (both in terms of average risk and in terms of changes in risk), and that we can then decide where the largest gains in improving road safety can be made in the (near) future.

These two ways of improving our understanding of the observed changes in road safety are certainly not mutually exclusive. In fact, they often complement each other. If we are interested in the effect of wearing a helmet

by (motor)cyclists on the number of fatalities, for example, then it would not make much sense to add the annual percentage of helmet wearing by (motor)cyclists as an explanatory variable to the modelling of the total developments in Figures 1.1 to 1.3. Clearly, in that case we at least need to disaggregate the total European Union road user population by road user type, and restrict the investigation to the effect of helmet wearing on the fatalities of (motor)cyclists only.

Deliverable D7.10 consists of three parts. In the first part (this part) a non-technical overview is given of the main results of all the time series analyses that were performed in the context of this document, while the more detailed description of the objectives, methods and results of each contribution can be found in Chapters 2 through 5 in the second part of the deliverable. In Chapter 2, the three important variables that can be distinguished in the analysis of road safety are discussed: exposure, accident risk and the consequences of the accident. In the other chapters, examples of applications of time series analyses to road safety are presented, where three ways of explaining the developments in road safety are used:

- disaggregation of the European data to national level to study the differences between the European countries (see Chapter 3)
- analysis of subgroups, by disaggregation by age, gender, traffic modes etc., to better understand the overall development in road safety (see Chapter 4 for a first step in this direction)
- analysis of the influence of a common external factor, such as weather conditions, law regulations or general changes in road infrastructure, on the road safety (see Chapter 5 for an example for weather conditions)

Part III of the report consists of a list of appendices where technical and mathematical details can be found, or large sets of graphical results are presented for the expert readers.

1.2 The three levels of risk

In the analysis of road safety, three important variables can be distinguished:

- Exposure (distance travelled, population size, fleet size etc.)
- Accident risk (annual number of accidents per annual distance travelled, per inhabitant, per passenger car).
- Consequences of the accident (number of fatalities or seriously injured per accident)

The study of the developments of these three levels of risk is not only interesting in itself. The distinction between the three risk level is also important when it comes to assessing *how* safety measures and other explanatory variables affect the total traffic process: by changing the exposure, the accident risk or the lethality risk, or two or all of these simultaneously. Time series analyses can be used to consider these three levels of risk simultaneously.

In Chapter 2, results of the multivariate analysis of these three levels of risk with structural time series models are presented as applied to quarterly data for the years 1987-2000 in France and the Netherlands, both at the national level, and stratified by road type for France. When comparing these results it

is found that the lethality risks of France and the Netherlands have had different developments in 1987-2000. Moreover, the lethality risk on French motorways decreased while it increased on French main roads.

1.3 Modelling road fatality trends in the European countries

The first obvious disaggregation of the development in Figures 1.1 to 1.3 that is interesting to investigate is the European Union *disaggregated by country*. Here we compare the developments in five of these countries: the Czech Republic, Spain, Belgium, the Netherlands, and Greece. More specifically, the annual number of road traffic fatalities per capita in each of these five countries was modelled in such a way that sudden changes in road safety developments can be detected. So, years when a previously increasing trend turns into a decreasing trend and vice versa, or where the rate of increase and decrease suddenly changes are modelled correctly. The results of these analyses plotted both against the independent variable motorization per capita and against time are shown in Figure 1.4.

Inspection of the middle panel in Figure 1.4 shows that the increase in the number of road fatalities per capita changed into a permanent decrease already in 1970 in the Netherlands, while in Belgium that turning point was reached in 1976. The main turning point in Spain occurred in 1989, and for Greece the turning point happened as late as in 1995. Countries like the Czech Republic and Spain exhibit more complicated patterns of breakpoints than the other four countries. In the Czech Republic, for example, turning points in the number of fatalities per capita are found in 1970 (from increasing to decreasing), 1988 (from decreasing to increasing) and around 1983 (from increasing to decreasing).

A most important feature of the top panel in Figure 1.4 is that they reveal for which number of motor vehicles per capita (that is, for which level of motorization rate) these breakpoints occurred. In Belgium the breakpoint is recorded by the time the motorization rate reached the value of 320 vehicles per 1000 inhabitants. For Greece the structural breakpoint is observed at a motorization rate of 335 vehicles per 1000 inhabitants. In Spain the main turning point occurred when motorization rate exceeded the threshold of 370 vehicles per 1000 inhabitants. In the case of the Netherlands the breakpoint corresponds to a value of motorization rate of around 210 vehicles per 1000 inhabitants, a relatively low value compared to the previous countries. In the Czech Republic the last turning point where the fatality risk started to decrease again is located at 320 vehicles per 1000 inhabitants.

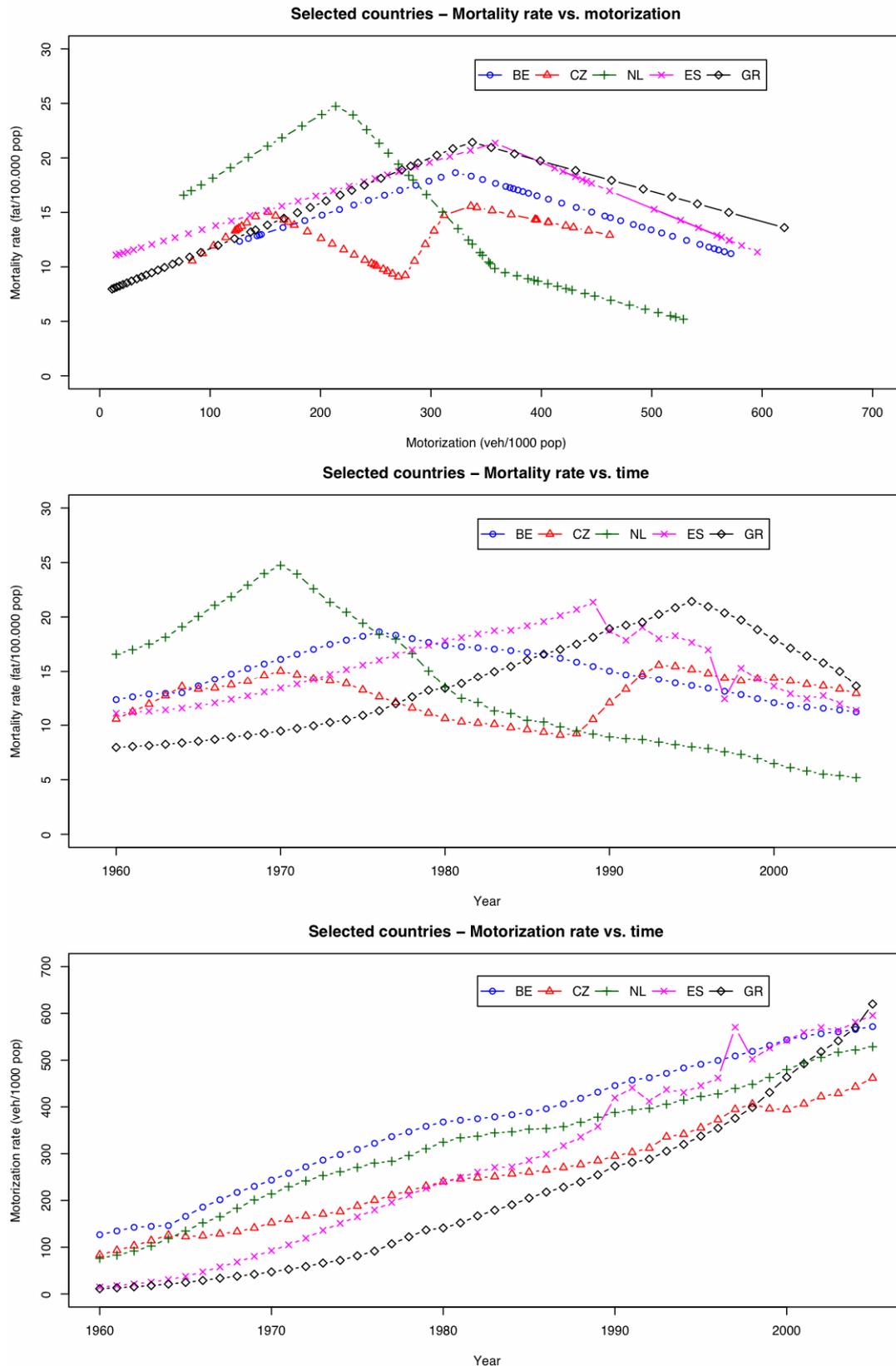


Figure 1.4. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

A most important feature of the top panel in Figure 1.4 is that they reveal for which number of motor vehicles per capita (that is, for which level of

motorization rate) these breakpoints occurred. In Belgium the breakpoint is recorded by the time the motorization rate reached the value of 320 vehicles per 1000 inhabitants. For Greece the structural breakpoint is observed at a motorization rate of 335 vehicles per 1000 inhabitants. In Spain the main turning point occurred when motorization rate exceeded the threshold of 370 vehicles per 1000 inhabitants. In the case of the Netherlands the breakpoint corresponds to a value of motorization rate of around 210 vehicles per 1000 inhabitants, a relatively low value compared to the previous countries. In the Czech Republic the last turning point where the fatality risk started to decrease again is located at 320 vehicles per 1000 inhabitants

This suggests that the major national turning points in the development of fatality risk occur within a certain specific range of the motorization rate, at least for some countries. In all cases there is some structural relation between the number of fatalities per capita (personal risk) and the number of motor vehicles per capita.

Summarising, we have seen that:

- different countries reach specific motorization rates in different years (temporal landmarks);
- some of these countries exhibit their major breakpoint in fatality risk within a narrow range of motorization rate values (320-370 vehicles per 1000), implying similar social and economical conditions;
- this range is different for certain subgroups of the examined countries, providing a hint that some grouping of countries may be required.

These preliminary findings should serve as an adequate base to pursuit further understanding of why and when such break points are observed. Research already conducted in this field will be taken into account so as to facilitate a useful grouping of the examined countries. The ultimate objective of next steps in the framework of this research should be to utilize this finding in order to make reliable predictions for countries or territories for which the so-called “break” has not occurred yet.

Disaggregation of the European data by country also clearly shows that different countries may show their own specific developments, and thus that disaggregation helps to understand the aggregated data. In the same way, disaggregation by other factors such as age, gender, traffic modes etc. can help us to better understand the road safety development at a national and European level.

1.4 Describing trends by disaggregation by age and gender

A more detailed level of disaggregation than discussed in the previous section involves the study of (future) developments in road safety while focussing on two demographic aspects of the road user population of a country: their age and their gender. It is well-known that young inexperienced car drivers are more at risk of being involved in traffic accidents than older more experienced

car drivers, for example; and in general male drivers are also more at risk than female drivers. All this implies that when the composition of a country's population changes over time this will affect the composition of that country's mobility on roads and therefore also the number of accidents and their outcomes.

One problem arising in this context is that –at least in the Netherlands- until recently annual mobility figures were only available for age *groups or classes*, while we would preferably have figures for each age *year* of the road user population. A second potential problem is that of possible *cohort effects* in the development of mobility of age groups, where the distance travelled by road users of 30 years old in 1985, say, was different from the distance travelled by road users of 30 years old in 2001, for example. Certainly, 50 year old drivers nowadays are far more experienced than 50 year old drivers in the seventies.

Information on the annual population size per age year of the Netherlands was used to break down the mobility figures for age classes into age years and, more importantly, the information on mobility figures and (forecasts of the) population size was combined to obtain smooth estimates of the Dutch mobility of car drivers stratified by age year both for the last 23 years and for the next 13 years (see Chapter 4 for details). The results are presented in Figure 1.5.

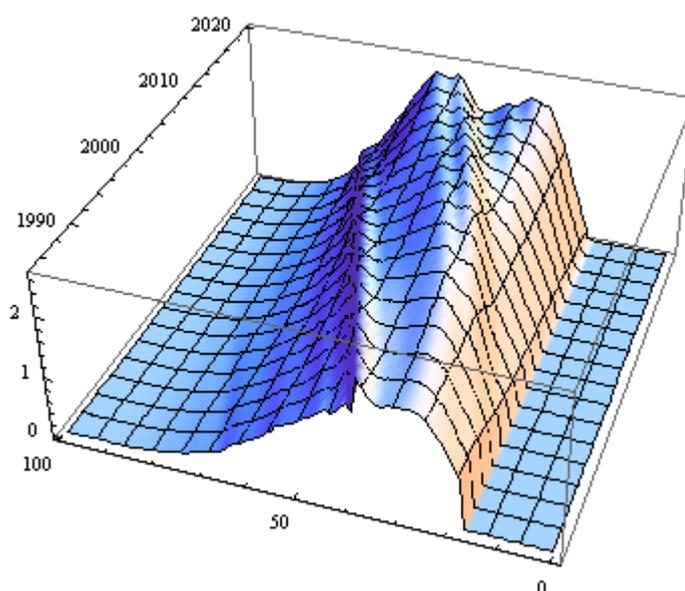


Figure 1.5. Estimated annual distance travelled (in billion km) by car drivers against their age year (from 0 to 100 years old) and time (front to back), including forecasts until the year 2020.

Figure 1.5 shows the (total) estimated annual distance travelled by car drivers, which consists of basically of two components:

- Population size
- Changes in mobility per capita

At this moment we focus on the demographic aspect. First of all, we see that the development of distance travelled by car drivers over time contains a 'wave' for car drivers of approximately 40 years old in 1985 extending

diagonally all the way to car drivers of approximately 70 to 80 years old in 2020. This wave can be identified as the post World War-II baby-boom of around 1946, as these car drivers were about 40 years old in the year 1985. Second, in the nineties a diagonal trough or valley starts to appear in the development of the mobility of car drivers that was almost continuously increasing until that year for car drivers older than about 25 years old. This could well be the effect of some social factors such as the introduction of the anti-conception pill for women around the end of the sixties and the individualization of the society in the seventies, as this reduction in children born around that time only emerges in the mobility figures of car drivers some 20 to 30 years later. It is finally interesting to note that these waves and troughs or valleys in the development of mobility run approximately parallel with the progressing age of the car drivers themselves. This suggests that there are no major cohort effects in the developments of the mobility of car drivers over time.

These results provided the impetus for deciding to use population forecasts (as issued by Statistics Netherlands) in order to obtain improved forecasts for future developments in mobility per age class and gender. The central idea is that, ultimately, these improved mobility forecasts are helpful in obtaining better forecasts for accident outcomes stratified by age and gender as well.

We therefore used three multivariate structural time series models to obtain forecasts of the mobility, the risk, and the number of fatalities and victims for the years 2008-2040 for each age class and gender combination of the total road user population (see Chapter 4 for all the technical details). In the first model (model 1) only figures on the distance travelled were used, in the second model (model 2) figures on the distance travelled and the population size were used, and in the last model (model 3) only figures on the population size were used.

In Figure 1.6, the forecasts obtained with these three models are compared for the fatalities disaggregated in four age classes based on data for the years 1985-2007, including intervention variables for the stratified risks in 2004. The latter intervention variables were added in order to capture the significant drop in the total number of fatalities observed in the Netherlands in the year 2004.

First of all, it is interesting to note that the male fatalities of 0-14, 15-24 and 25-64 years old and the female fatalities of 15-24 years display significant reductions in the year 2004, but that these reductions do not apply to the other subgroups.

More importantly in the present context, we see that the inclusion of population figures in the analyses is especially relevant for the forecasts of male fatalities in the 15-24, 25-64 and 65+ age groups, and for the forecasts of the female fatalities in the 25-64 and 65+ age groups, since the forecasts obtained with the three models diverge over time for these groups. Traffic safety forecasts based on fatalities and mobility data alone, can therefore be improved by disaggregation of these data by age, and inclusion of the expected trend of the composition of the population.

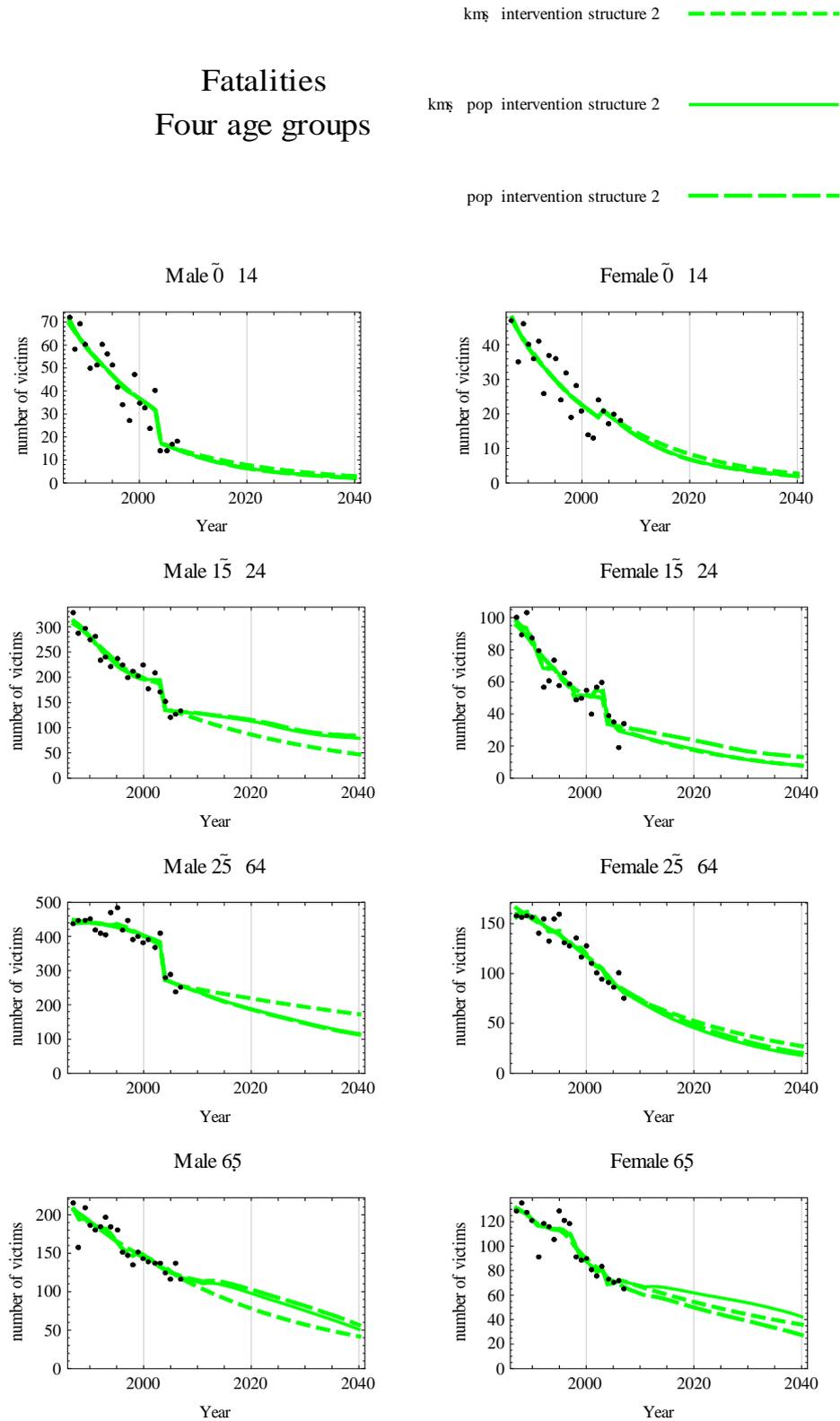


Figure 1.6. Forecasts from 2007 onwards of fatalities in 4 age classes as obtained with mobility figures only (model 1), combined mobility and population figures (model 2), and population figures only (model 3) including intervention variables to capture the significant drop in the total number of fatalities in the Netherlands in 2004.

For some subgroups at least, the inclusion of (forecasts of) population figures in these analyses clearly allow to better take the distinct waves and troughs or valleys in the stratified mobility developments observed in Figure 5 into account. Since the development in the age composition of the general population of a country affects the development of the mobility of age years and groups of the road user population, the incorporation of these waves and troughs extending into the future seem extremely relevant for forecasting mobility figures stratified by age and gender, and are therefore helpful in obtaining improved forecasts for road accident outcomes as well.

1.5 Modelling meteorological effects on road safety

At first sight it may seem that the investigation of the relationship between weather conditions and the occurrence of road accidents and their consequences is a futile exercise, for the obvious reason that there is no way in which we can exert any influence on these weather conditions. The importance of studying these relationships immediately becomes apparent, however, as soon as it is realized that the study of the effects of factors on road safety that we indeed can control are often obscured by such confounding factors as weather conditions. The ultimate message is that – whenever possible- those factors influencing road safety that are outside our control should always be taken into account when studying the effects of measures for the improvement of road safety over which we do exert a certain amount of control.

According to relevant literature, weather may explain in the order of 5% of the accident/fatality variability. Therefore, models that include only weather effects as explanatory variables cannot be expected to have predictive value. Instead, these models can be expected to provide some insight into the expected effects of weather conditions both in the form of type of effect (positive/negative) and relative magnitude. In the first subsection results are presented for the Athens region using temperature and precipitation data on a monthly as well as daily level. The main reason was that the exposure data were only available at a monthly basis, but the weather and accident data were available at a daily level. In the second subsection, the effects of temperature, rainfall and frost were investigated for France, the Netherlands and the Athens region, using aggregated data at the monthly and national level. There were several differences in approach: in subsection 1.5.1 mean temperature of the day was used and the number of accidents and fatalities was the outcome measure, while in subsection 1.5.2 maximum temperature of the day was used and the outcome measure was the number of injury accidents. Nevertheless, the two approaches yielded to coherent results, when used on the same dataset: the monthly number of injury accidents in the Athens Region.

1.5.1 The Athens region

To investigate the effect of two essential meteorological indicators (temperature and precipitation) on the number of total accidents and fatalities recorded in the wider Athens area for the period 1985-2005, daily average temperature and total precipitation for the wider Athens area were used.

Furthermore, in an attempt to also consider exposure data, monthly traffic data have been used for the period from January 1997 to December 2005 (108 months), as collected at the Shimatari toll station north of Athens. These data are assumed to be a reasonable approximation of the entire traffic in the Athens area. Information on the traffic composition (passenger cars etc) was also available and is exploited in the analyses.

To be able to meaningfully incorporate the exposure data at a monthly level (only available from 1997 onwards) in the analysis, the meteorological data were also (a) truncated to only include data after 1997 and (b) aggregated to the monthly level. For the aggregation of the precipitation, a total precipitation for the month was computed. For the aggregation of the temperature, several variables were computed: month average mean temperature, minimum mean temperature for the month, and maximum mean temperature for the month (where mean refers to the daily mean). Figure 1.8 displays the thus aggregated monthly data. There is evidence in the literature that meteorological variables may be more useful if they are categorized, allowing for the evaluation of the effect of extreme temperatures or days with high precipitation. The continuous values of the variables temperature and precipitation were therefore classified in just five and seven categories, respectively.

Several observations can be made concerning Figure 1.7. Unusually high precipitation was mainly recorded during the winter of 2002-03, following two relatively dry winter periods. Moreover, the data reveal a significant reduction of fatalities and total accidents. It is stressed, though, that this reduction practically appears as a step in the transition from 2001 to 2002. On the other hand, a steady increase of passenger car traffic is recorded at the annual level (except for a big drop in February 2002).

The proportion of passenger cars travelling through the aforementioned toll station appears to increase over time. In particular, cars amounted up to approximately 75% of all vehicles for the first three years (1997-99). This proportion increased to 80% in 2000, 85% for the period 2001-2004 and about 90% in the second half of year 2005.

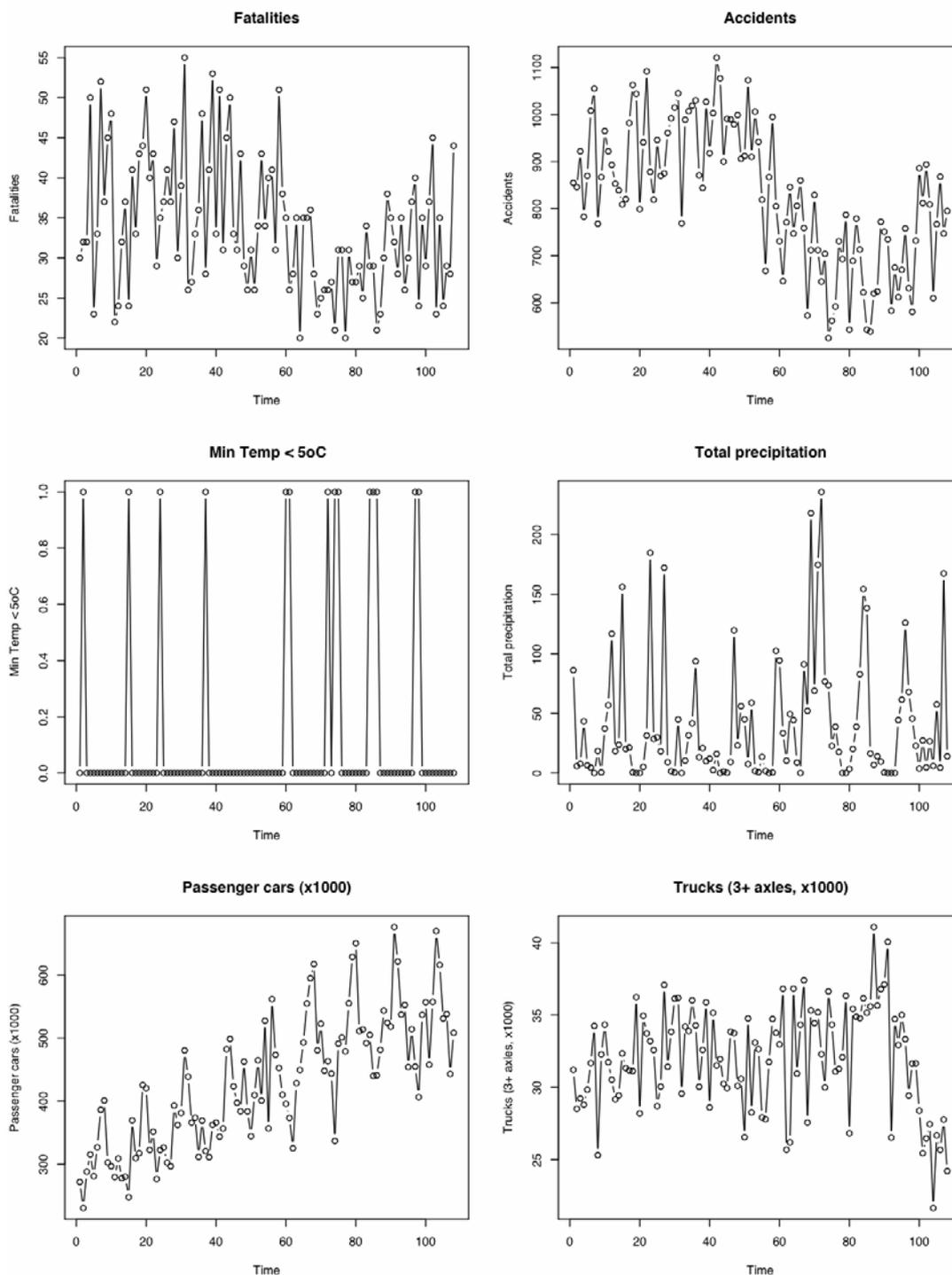


Figure 1.7. Descriptive graphs of the monthly dataset used for the analysis

Working with different data sources inevitably leads to mismatches in the granularity and scope of the data. In this case, meteorological and accident data are available at a much higher resolution (i.e. daily) and span a much longer time period (1985 to 2005) than the exposure data which are only available at a monthly level, and then only for the last 9 years. In order to both explore the detailed meteorological data and obtain some insight into the effects of traffic, two different analyses have been performed: one with the meteorological data alone (using daily time series data between 1985 and

2005) and one with the aggregated time series (used monthly data from 1997 to 2005). Separate generalised linear models were used to predict the accident and the fatality data using a negative binomial distribution assumption on the residuals.

For the daily accident data it is found that the increase in temperature results in an increase in the number of accidents. An increase in precipitation, on the other hand, is associated with a decrease in the number of accidents, possibly due to extra care exercised by the drivers. When comparing the months of the year, smaller numbers of accidents are observed in the month of August, reflecting the fact that smaller traffic counts are observed during that month in the Athens region.

For the daily data on fatalities higher temperatures are also associated with larger numbers of fatalities while very high precipitation values are associated with a decrease in the number of fatalities.

When monthly data are considered -including the effects of exposure- it is found that very low temperatures (that is, months with minimum daily temperatures below 5° Celsius) are associated with a reduction in the number of accidents, while higher monthly total rain precipitation results in smaller numbers of accident. Furthermore, higher passenger car traffic results in a lower accident number (presumably due to the lower speeds that are allowed due to higher congestion), while larger numbers of heavy trucks leads to an increase in accidents. Similar relations are found in the analysis of road fatalities at the monthly level.

1.5.2 Modelling the weather effect in different regions

The effects of weather conditions on road safety were also evaluated both at the national level and for different types roads for France and the Netherlands, and for the same region of Athens as discussed in the previous subsection. Not only the effects of rainfall and temperature were investigated, but also those of frost. For France and the Netherlands, the daily observations of these three weather variables from different meteorological stations were first averaged to obtain data at the national level. For the region of Athens, France and the Netherlands they were then aggregated at the monthly level.

At the national level it is found that the global weather effects on the number of injury accidents are very similar for France and for the Netherlands:

- rainfall is positively correlated with the number of injury accidents
- temperature is positively correlated with the number of injury accidents (100 mm of additional rainfall during a month results in an increase of 0.2%-0.3% in the number of injury accidents of the month)
- the occurrence of frost is negatively correlated with the number of injury accidents (1 additional day of frost in the month results in a decrease of 0.3%-0.6% in the number of injury accidents of the month)

However, on certain types of road the weather effects can be different from those found at the national level, in magnitude and sometimes even in sign:

- rainfall is still positively correlated with the number of injury accidents. But in France a 100 mm increase in rainfall in the month results in an increase of about 0.5% in the number of injury accidents on main roads and on motorways, while the effect in the Netherlands is an increase of 0.9% in the number of victims on motorways, and not significant on rural and urban roads. Moreover, in the region of Athens rainfall is even negatively correlated with the number of injury accidents: there 100 mm additional rainfall in the month results in a decrease of about 0.5 % in the number of injury accidents;
- temperature is positively correlated with the number of injury accidents, but its magnitude depends on the considered road type for France and the Netherlands. Moreover, in the region of Athens, however, the impact of temperature is only significant in the winter.
- the occurrence of frost also has different effects depending on the type of road. For example, one additional day of frost per month results in a decrease of 1% in the number of victims on urban roads in the Netherlands.

In Section 5.3, also some results are presented about the decomposition of the global weather effects into direct effects (i.e. directly on the number of injury accidents) and indirect effects (i.e. effects through changes of mobility). The most significant result is related to rain. On main roads and motorways in France, an increase of rainfall is associated with a decrease of the distance travelled, and with an increase of the number of accidents under constant mobility conditions.

Part II

Chapter 2 - Modelling the three levels of the road risk process at an aggregate level

Ruth Bergel, Mohammed Debbarh (INRETS), Frits Bijleveld, Jacques Commandeur, Henk Stipdonk (SWOV)

2.1 Introduction

The road safety level is usually represented by the total number of (fatal or serious) accidents, or casualties. Therefore it is not surprising that analyses or forecasts of traffic safety concentrated on the aggregate *numbers of injury accidents* (mainly injury accidents with killed and seriously injured persons) *and victims* (mainly fatalities) on the road. These analyses and forecasts were initially carried out as annual time series analysis, but later extended to quarterly, monthly or even daily time series analyses. The final goal of these analyses is to understand and predict the development of road safety, and the effect of important factors such as safety measures on road safety.

In France, descriptive analyses were performed by the Road Safety Observatory since 1972, the year this observatory was created. These data show seasonal variations and large fluctuations due to chance. Such variations obstruct the description and explanation of *the trends* of the road safety indicators in an appropriate manner, either for short term or a long term (ONISR, 2006). Moreover, the relationship between the road safety indicators and the national safety measures was questioned, and a systemic approach was adopted in order to take into account all types of risk factors. Among these factors, exposure to risk (e.g. distance travelled) plays a most relevant role, so efforts were made to take the effect of the *exposure to risk* into account.

Different types of variables can be used to measure exposure to risk: the population size, the vehicle fleet, the number of persons, or the distance travelled by persons or vehicles (expressed in person-km or vehicle km). Proxies such as the oil sales, on a monthly basis were also used. A state of the art of the existing exposure data and methodologies was produced within the SafetyNet project (Yannis et al; 2005). The methodological framework adopted in road safety research has already been presented in the SafetyNet Deliverable D7.4, so for precise definitions of some major concepts related to the road risk we also refer to that Deliverable.

In the analysis of road safety, three variables are distinguished:

- A measure of exposure (distance travelled, population size, fleet size etc.)
- An accident's risk indicator (the number of injury accidents or a risk rate, such as the number of injury accidents divided by the related exposure measure).

- an accident's severity indicator (the number of victims or a severity rate, such as the number of victims divided by the related number of injury accidents or directly divided by the related exposure measure)

The relevance of considering *the road risk process in its three main variables (sometimes also called levels)* has been highlighted by road safety researchers, and is now adopted at an international level (Lassarre, 1984). Within that framework, a time series analysis is to be performed on these three indicators altogether.

The most straightforward and simple approach to analyse road safety is to jointly monitor annual exposure to risk, annual number of accidents and annual number of victims (or alternatively monthly data). Another approach is to directly monitor exposure, and the respective ratios between exposure and accidents, and between accidents and victims, instead of the absolute numbers of accidents and victims. This approach is retained in this chapter, and has the advantage of highlighting the development of the road risk process' structure itself. The accident risk and the severity risk are the two aggregate road safety indicators of major interest.

Comparison of road safety data of European countries should ideally be performed on the basis of risk ratios such as the number of injury accidents and victims registered for a whole country reported to an *exposure to risk measure*. In Chapter 3, for instance the *mortality* (fatalities divided by the number of inhabitants) is used to compare several European countries. However, mobility (distance travelled) is the preferred exposition measure, to enable comparison of accidents per distance travelled, (Yannis et al., 2005, SafetyNet Deliverable 2.1, and related recommendations, forthcoming 2008). Lassarre (2001) already demonstrated the importance of such an indicator for modelling the annual number of fatalities for ten European countries.

The aim of this chapter is to provide some examples of time series analysis, applied to road safety data aggregated at a national level, and *disaggregated according to a network or to an accident and victims typology*. The framework is of a *descriptive approach*, as defined above. An aggregated measure of exposure to risk (whether the distance travelled on a road network, or a proxy such as the fuel consumption related to the network, for instance) is used, and we shall therefore analyse simultaneously the development over time of the three levels of the road risk process' indicators, which were defined above. The period chosen for the analysis is 1987-2000, as it includes a turnover in the fatality risk trend for both countries.

2.2 Method

The analysis of road safety data requires some technical tool, such as regression of a model against the exposition and risk data. In the SafetyNet analysis, the structural time series methodology or state space analysis is applied (Harvey, 1989), (Durbin, Koopman, 2001), (Commandeur, Koopman, 2007). The econometric specification used for the structural time series methodology is recalled in Appendix A2.1.

In all the applications presented in this Chapter, the three observed variables mentioned in the former paragraph:

- a measure of exposure
- the number of injury accidents
- the number of fatalities

are modelled using three datasets defined for the same time period. For these three variables, the analysis uses the actual data to estimate the most likely value of the unobserved *expected* values for:

- exposure
- injury accidents per distance travelled
- number of fatalities per accident.

These unobserved values are chosen such that the likelihood of the data, given the unobserved values, is optimum.

The final aim here is to estimate the unobserved *components* (the *level* and *slope* of a locally linear trend, the *seasonals*, and the *residuals*) of the expected values of the three variables. The importance of these components lies in the development of their values over time.

An important advantage of analysing exposition, number of accidents and severity at the same time, is that this allows for capturing the correlation between the observed variables. Separate analysis would not show a possible correlation between the development over time of e.g. exposition and accidents. Preceding references for a structural time series analysis applied to road safety data at a national level, in the univariate and bivariate cases, are (Bergel, 1997), (Lassarre, 2001) and (Cost 329, 2004)

The structural time series analysis has several important advantages over classical regression methods, most of which are of a technical nature. Besides, it enables an accurate picture of the change over time from year to year or even from month to month, of the hidden development of risk over time.

2.3 Data

The data, on a quarterly basis and for the period 1987-2000, used for France were:

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

and for the Netherlands:

- the vehicle kilometres of drivers of all vehicles for the whole of the Netherlands, based on survey data,
- the total number of injury accidents (resulting in people being killed or seriously injured),
- the total number of people being killed.

The data used, disaggregated according to a typology, were for France:

- the number of vehicle kilometres,
- the total number of injury accidents,
- the total number of people being killed, registered on main roads, and on motorways,

and for the Netherlands:

- the number of vehicle kilometres,
- the number of KSI accidents,
- the number of people killed or seriously injured,
- the number of fatalities, registered by KSI accident type.

Note that the absolute values of the data are not strictly comparable between the two countries, as the definitions of the indicators is in no way comparable. Nevertheless, their development over time is comparable, and that comparison makes sense, as it will be developed later on.

2.4 Results

Model Fit and Diagnostics

For each variable, the level, slope, seasonal component and residuals were estimated simultaneously, and the variances of their disturbances provided (see tables A2.1 and A2.2 in Appendix A2.1).s The usual tests of validity were performed on the residuals (independence, homoscedasticity and normality), and were satisfactory.

Main results

Figures 2.1 to 2.4 show the development of the data and components related to France and The Netherlands the period 1987-2000, on a quarterly basis. The logarithms of the values are plotted. For every estimated value, a 95% confidence interval is given. This indicates there is 95% probability that the actual value lies within this interval.

France

Figures 2.1 and 2.2 show the development of the data related to France. The expected values of the three observed variables (oil sales as a measure of exposition, the number of injury accidents and the number of fatalities), estimated with the model, are recalled in Figure 2.2. The seasonal pattern for each variable appears to be different, but rather stable throughout time. The trend of the oil sales shows a continuous increase, whereas the trend of the number of injury accidents and fatalities shows a continuous decrease. But the trend, especially in the case of the two last indicators, appears not to be linear.

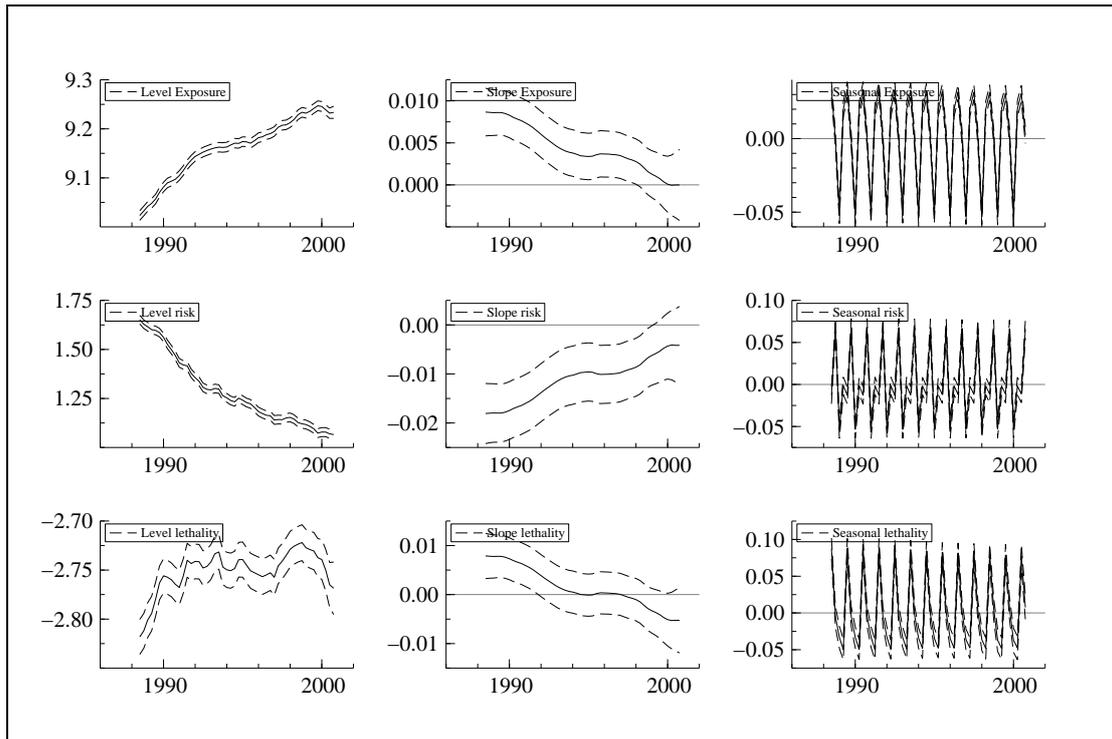


Figure 2.1: Unobserved components: levels, slopes and seasonals, (from left to right) for exposure, accident risk and fatality risk (from top to bottom) for France. The estimated values are given with their 95% confidence intervals.

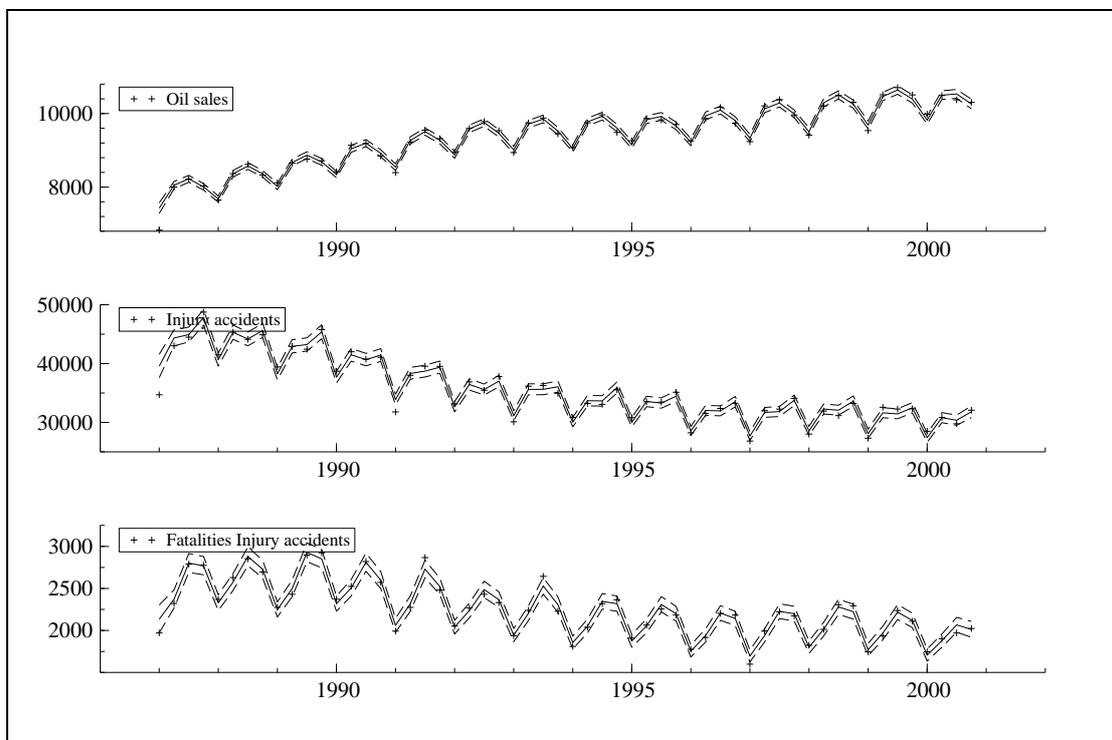


Figure 2.2: Estimated and observed data, where the estimates are based on the results in figure 2.1 (1987-2000). Quarterly oil sales (top panel), injury accidents (middle panel), and fatalities (bottom panel) for France. The estimated values are given with their 95% confidence intervals.

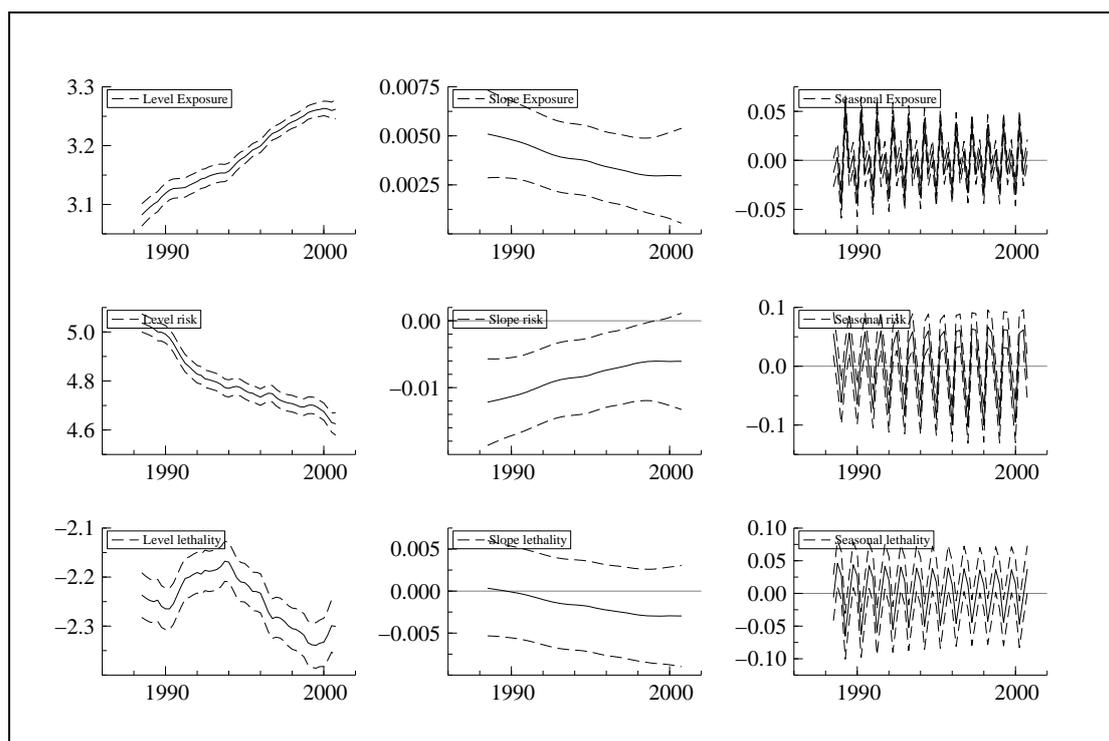


Figure 2.3: Unobserved components: levels, slopes and seasonals, (from left to right) for exposure, accident risk and fatality risk (from top to bottom) for The Netherlands. The estimated values are given with their 95% confidence intervals.

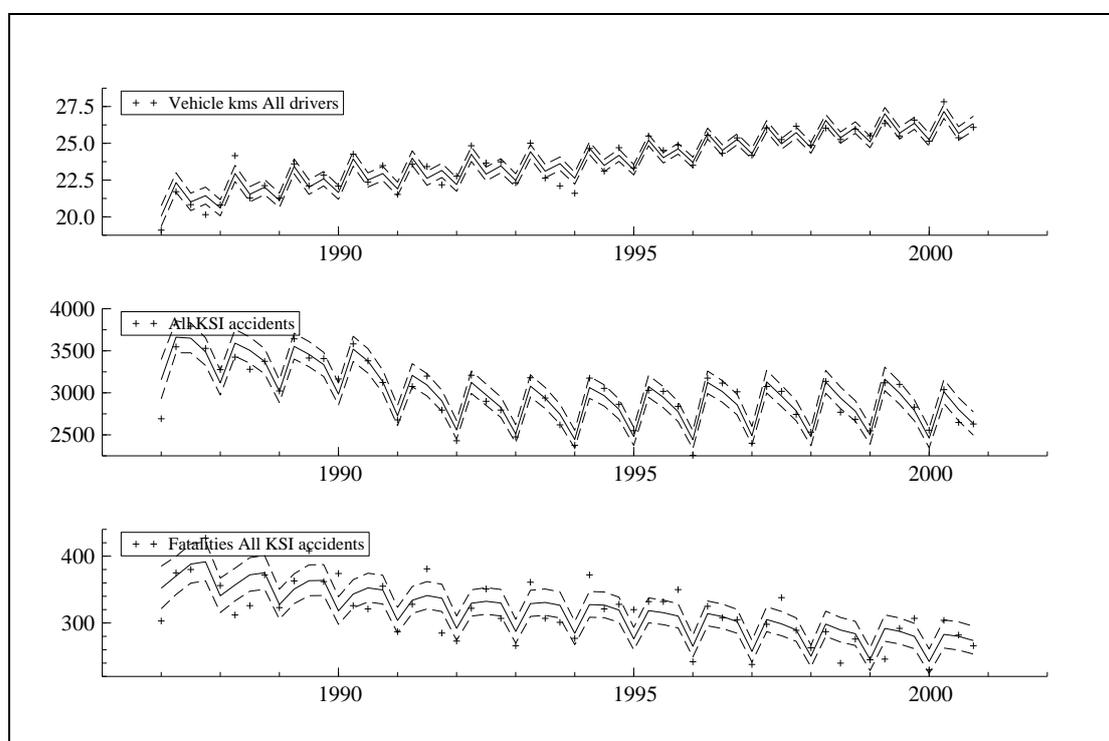


Figure 2.4: Estimated and observed data, where the estimates are based on the results in figure 2.1 (1987-2000). Quarterly oil sales (top panel), injury accidents (middle panel), and fatalities (bottom panel) for the Netherlands. The estimated values are given with their 95% confidence intervals.

The results presented in Figure 2.2 are based on the calculated unobserved level (left column), slope (middle column) and seasonal (right column) components of exposition, accident risk and fatality risk, presented in Figure 2.1. The exposure in France has continuously increased. This follows from the second graph in the first row of Figure 2.2, where the time-varying slope component for exposure is positive throughout the period 1987-2000. Moreover, since the positive value of the slope becomes smaller towards the end of the period, the rate of increase in exposure has decreased since the beginning of this period. Similarly, inspection of the second graph in the second row of Figure 2.2 shows that the slope component for accident risk is negative throughout the time period, but that its absolute value becomes smaller and smaller as time proceeds. At the end of the period the increase may even have come to a halt. This means that the accident risk in France has continuously decreased, but that the rate of decrease is levelling off towards the end of the period. Finally, the fatality risk has first increased until about 1994, but has started to decrease after 1997, as follows from inspection of the second graph in the bottom row of Figure 2.1.

As for the seasonal components in the last column of Figure 2.1, in France the exposure and the fatality risk are always the largest in the third quarter of the year, while the accident risk is always the largest during the last quarter of the year.

The Netherlands

Just as in France, the exposure in the Netherlands, corrected for seasonal effects, has continuously increased, while the seasonally corrected accident risk has continuously decreased. This follows from the values of the corresponding slope components, which can be seen in Figure 2.3, top left panel. In the Netherlands the fatality risk has continuously decreased from 1990 onwards (Figure 2.4, bottom middle panel, where we see that the slope is negative in the entire period), while in France it first increased until about 1994, only starting to decrease after 1997 (Figure 2.2, bottom middle panel, where we see that the slope is positive until 1994, and negative after 1997).

Inspection of the confidence intervals of the slope components in Figure 2.3 shows that both the upper and lower bounds for exposure are positive. This means there is little doubt about the increasing exposure. Also, the upper and lower bounds for accident risk are all negative. This means there is little doubt that this accident risk really decreased. However, the upper bound of the confidence interval of each individual slope value for fatality risk is consequently located on the positive side of the horizontal axis while the lower bound is consequently located on the negative side of the horizontal axis. This means that the sign of the development in the slope of the fatality risk in the Netherlands is more uncertain. Actually, we cannot rule out the possibility that the fatality risk did not actually change much over time.

The seasonal components in Figure 2.4, for The Netherlands shows that exposure is largest for the second quarter of the year, accident risk is largest during the third quarter of the year (although the second quarter becomes

more and more similar to the third quarter towards the end of the period), while fatality risk is largest during the last quarter of the year.

Comparison between France and the Netherlands

It is not possible to compare the developments of the three levels for the Netherlands and France in absolute terms. For example, the range on the y-axis of the trend of the unobserved fatality risk is between $\exp(-2.4) = 0.091$ and $\exp(-2.1) = 0.122$ for the Netherlands in Figure 2.3, suggesting a range of 0.09 to 0.12 persons killed per injury accident in this country (9-12 persons killed for 100 injury accidents). For France, on the other hand, the range on the y-axis of the trend of the unobserved fatality risk in Figure 2.1 is between $\exp(-2.85) = 0.059$ and $\exp(-2.7) = 0.067$, suggesting a range of only 0.06 to 0.07 persons killed per injury accident (6-7 persons killed for 100 injury accidents), and therefore a smaller accident severity than for the Netherlands. These smaller numbers are however at least partly explained by the different definitions used in the two analyses for an injury accident, the definition for France being much more liberal (and therefore involving many more, less severe accidents) than for the Netherlands. However, the use of a correcting factor applied to the number of fatalities as defined in France before 2005 would allow correcting for that difference to some extent.

Main roads and motorways in France

The part of the French main network that is measured by sensors, carries about one third of the mobility (distance travelled) in France, and around 14% of injury accidents and 28% of fatalities (data for the year 1998), as shown in Table 2.1. Although the mobility is equally shared between main roads and motorways, this is not the case for the number of injury accidents and the number of fatalities, and as a consequence for the accident risk and fatality risk on each network. Moreover accident risk and fatality risk developments for the period 1997-2000 show differences between the two networks, and with France.

	Main roads	Motorways	Main network	Secondary roads	Urban roads	France
Injury accidents	11 807	5 910	17 717	36 278	70 392	124 387
monthly average	984	493	1 476	3023	5866	10 366
%	9,5%	4,8%	14,2%	29,2%	56,6%	
Fatalities	1 928	471	2 399	4 373	1 665	8 437
monthly average	161	39	200	364	139	703
%	22,9%	5,6%	28,4%	51,8%	19,7%	
Mobility 10⁹ veh km	88,6	96,9	1 85,5			
monthly average	7,4	8,1	15,5			
%			ca 30%			
Network length (km)	24 000	8 763	32 763			

Table 2.1: Traffic and accident data for the year 1998 for France

Figure 2.5 to 2.8 show data on exposure and accidents on French main roads and French motorways respectively. The mobility on French main roads is described quite accurately, as reflected in the fact that the 95% confidence limits for the estimations for this variable in Figure 2.6 are now much smaller than the confidence limits for oil sales in Figure 2.2.

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

Again we see that the distance travelled on French motorways is described quite accurately since the 95% confidence limits of the estimations for this variable in Figure 2.8 are now also much smaller than for the oil sales used as a proxy for exposure in the analysis for the entire French network in Figure 2.2.

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

2.5 Discussion and perspectives

Both in France and in the Netherlands, exposure continuously increased and the accident risk continuously decreased. However, the fatality risk did not develop in a monotonous way, and differences appear:

- In France, the fatality risk increased until about 1994 and started to decrease after 1997 (it decreased on the motorways network but it increased on the main roads network)
- In the Netherlands, the fatality risk continuously decreased from 1990 onwards

These differences highlight the interest of considering disaggregated datasets according to the network category, or even other disaggregations, as described in Chapter 4.

The perspectives of this approach first consist in the constitution of an exposure to risk measure, such as the number of kilometres or persons-driven at a national level in Europe (Yannis et al, 2005 and on-going recommendations).

In the absence of such a measure - or, if not, of an estimation - the descriptive analysis of the road risk process at a national level will continue to be performed on the two levels of the road risk process, measured as it is commonly the case with the aggregate numbers of injury accidents and fatalities. Some examples of this two-level analysis, including explanatory variables, are given in chapter 5.

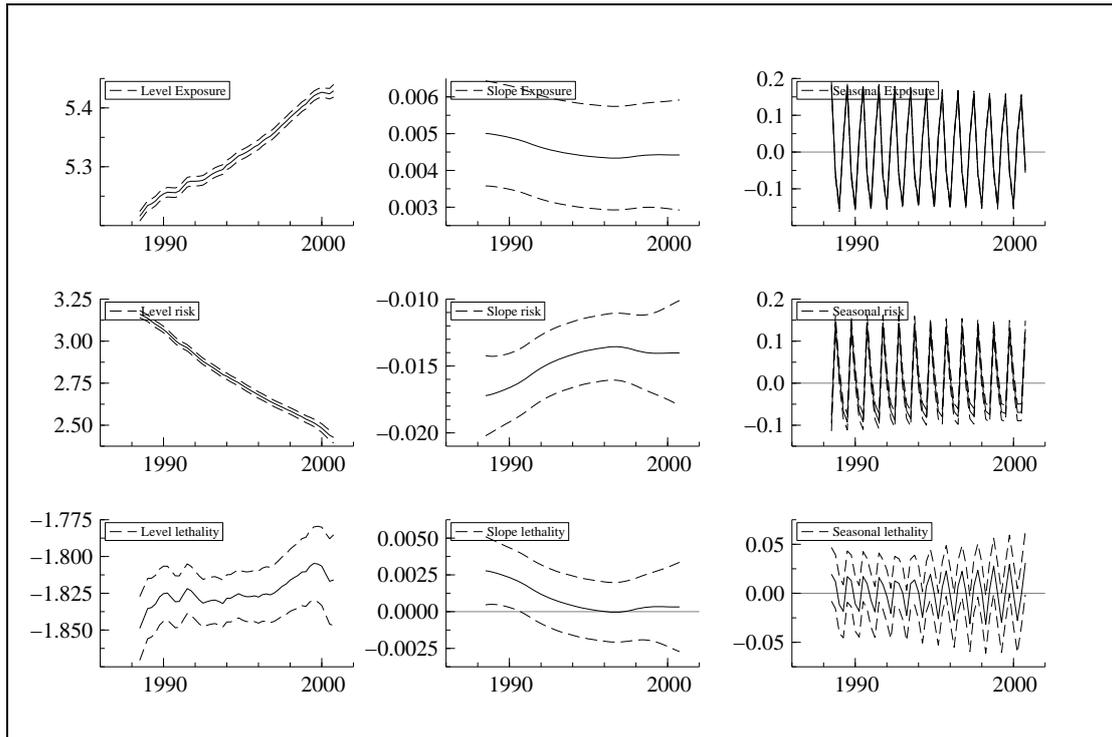


Figure 2.5: Unobserved components: levels, slopes and seasonals, (from left to right) for distance travelled, accident risk and fatality risk (from top to bottom) for main roads in France. The estimated values are given with their 95% confidence intervals.

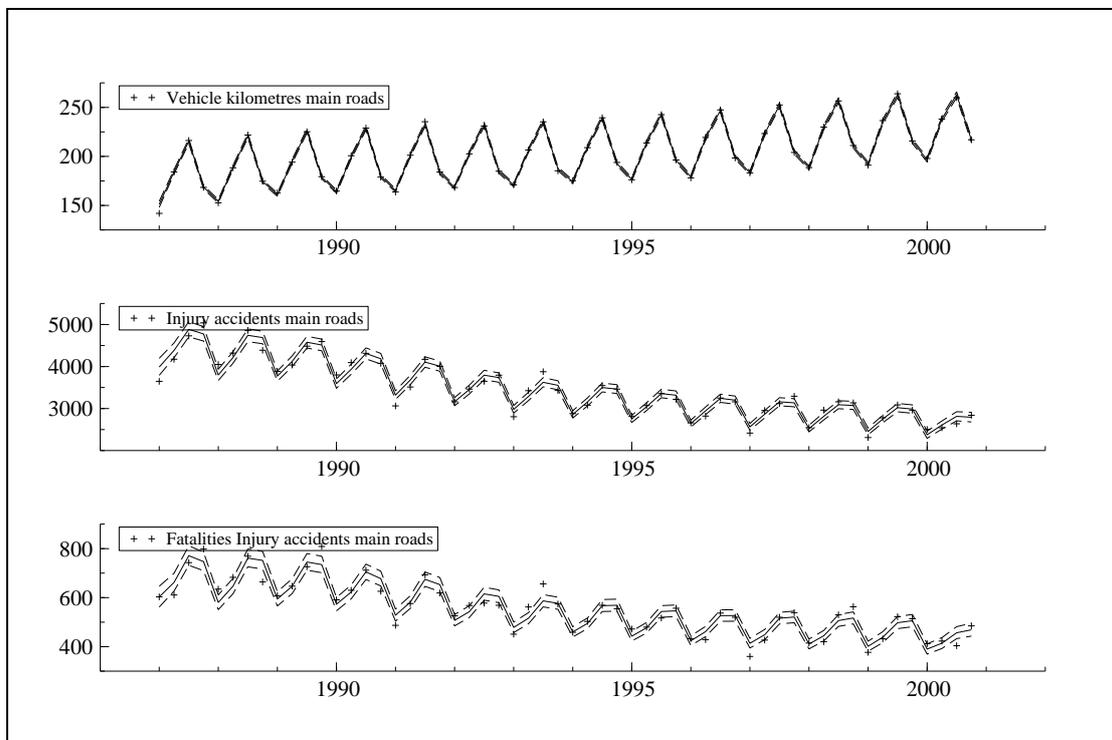


Figure 2.6: Estimated and observed data, where the estimates are based on the results in figure 2.1 (1987-2000). Quarterly distance travelled (top panel), injury accidents (middle panel), and fatalities (bottom panel) for french main roads. The estimated values are given with their 95% confidence intervals.

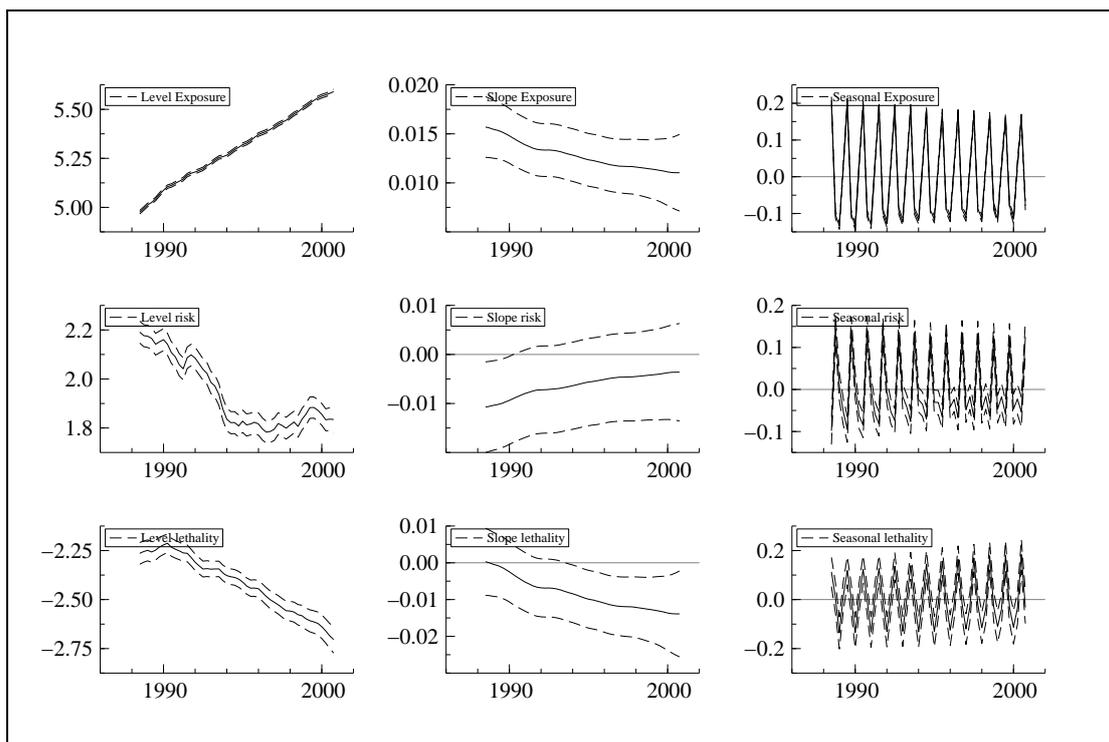


Figure 2.7: Unobserved components: levels, slopes and seasonals, (from left to right) for distance travelled, accident risk and fatality risk (from top to bottom) for french motorways. The estimated values are given with their 95% confidence intervals.

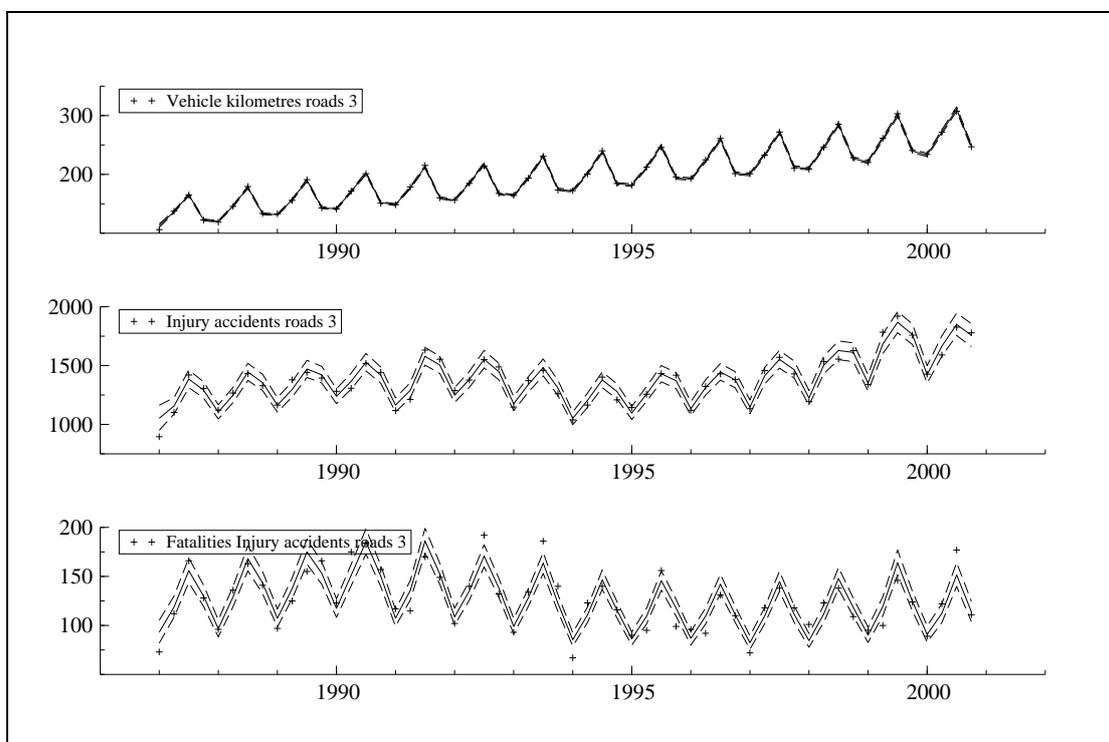


Figure 2.8: Estimated and observed data, where the estimates are based on the results in figure 2.1 (1987-2000). Quarterly distance travelled (top panel), injury accidents (middle panel), and fatalities (bottom panel) for french motorways. The estimated values are given with their 95% confidence intervals.

2.6 References

Bergel R. (1992). *Multivariate modelling of road risk on the French main road network*. Proceedings of the 25th European Transport Annual Meeting PTRC, Brunel University, Uxbridge, 1-5 September 1997,

Bijleveld F.D., Commandeur J.J.F., Gould P.G., Koopman S.J. (forthcoming 2008). Model-based measurement of latent risk models in time series with applications. *Journal of the Royal Statistical Society Series A*.

Commandeur J., Bijleveld F., Bergel R. (2007). *A multivariate time series analysis applied to SafetyNet data*. SafetyNet Deliverable D7.7, <http://www.erso.eu>

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

Commandeur J., Koopman S.J. (2007). *An introduction to state space time series analysis*, Oxford University Press.

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

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

Observatoire National Interministériel de Sécurité Routière *La sécurité routière en France : bilan de l'année* (2006). ONISR, Ministère de l'Équipement, La Défense.

Lassarre S. (1994). *Cadrage méthodologique d'une modélisation pour un suivi de l'insécurité routière*, Synthèse INRETS, n°26, Arcueil.

Lassarre S. (2001). Analysis of progress in road safety in ten European countries. *Accident Analysis & Prevention*. Vol. 33, pp. 743-751.

Yannis G et al. (2005). *State of the Art Report on Risk and Exposure Data*. SafetyNet Deliverable 2.1 <http://www.erso.eu>

Chapter 3 - Modelling road fatality trends in the European countries

George Yannis, Constantinos Antoniou, Dimitris Katsochis, Eleonora Papadimitriou (NTUA)

3.1 Introduction

The comparative analysis of macroscopic trends in road-safety-related issues for EU countries has been the favourite topic and specialization area of many researchers. Since 1949, when R.J. Smeed first presented his approach on the development of traffic safety in a large number of countries showing a relation between the number of fatalities, vehicles and population size, a lot of interesting research work has been carried out.

Smeed's contribution in opening research towards macroscopic indicators trends and comparative analysis of road safety level at international level has been undoubtedly essential. His initial approach that entailed data for twenty (mostly European) countries from only one year (1938) yielded traffic risk (fatality rate) as a simple (exponential) regression function of the motorisation rate. This was followed by pieces of work utilising datasets over periods of time (Smeed, 1968 – elementary formula on the accident rate patterns across countries). As data complied with this approach until mid-seventies, Smeed's formula gained overall acceptance, often referred to as "Smeed's law".

Of course, there have also been over time several researchers that criticised Smeed's work as lacking adequate theoretical background. It is true that the formulae in question have been limited in terms of real life interpretation up to the point that fatalities per head of population started to decrease. From 1975 onwards, this approach also lacked its empirical validity. Still, this seemingly empirical attempt is important and influential indeed, since it served as a reliable tool of depiction through early stages of road safety developments (COST, 2004). Furthermore, it cultivated the expectation that time series at international, macroscopic level may well provide valuable insight on each nation's road users' capability of learning from their past experiences.

Smeed mainly addressed the issue of illustrating past trends. His models mainly fall in the so-called "first-generation" category of cross-sectional models (product benchmarking level). According to this classification (see e.g. Al Haji, 2007), the second generation of models was formed with the introduction of the time dimension, proceeding from descriptive to explanatory. Finally, two more generations have been introduced by means of integration between the product (fatalities rates) and other model indicators (e.g. related policy strategies) resulting to foreseeing models.

In recent years, some most interesting approaches have been presented on the topic. Lassarre (2000) proceeded to an analysis of ten European countries progress in road safety by means of a structural (local linear trend) model,

yielding two adjusted trends, one deterministic and one stochastic. This was achieved through incorporation of intervention functions related to the major road safety measures introduced. An indicator of the rate of progress given risk exposure trends (vehicle-km travelled) was defined. It was deduced that all major EU countries exhibit rate of progress above a minimum threshold of 4.5% annually. The average of 6% is equivalent to the statement that the road transport system in Europe “is capable of absorbing a 6% increase in traffic at a constant number of fatalities”.

Page (1997) presented an exponential formula that yields fatalities as the product of all explanatory variables' influence. This function is easily transformed to a simple algebraic form (first order polynomial with an intercept) by taking the logarithm of both sides. The objective of the respective paper lays on safety level comparison among selected OECD countries during the period 1980-1994. Taken that fatality rates are not sufficient to perform international comparisons, a statistical multiple regression tool was set up. The model assumes that each variable's effect is independent of its original level, implying constant elasticity. Interestingly, the negative intercept represents a positive mean effect of missing variables on safety.

Other analyses entail a specific road safety related problem, applying international macroscopic comparison techniques to a subset of road network users, such as novice or young drivers. Twisk & Stacey (2007) presented a general study of identified trends in young drivers risk and associated countermeasures in certain European countries. The relationship between general safety levels and young driver risks is stressed: the impact of general safety measures on the subgroup is greater than that of measures specifically targeting young drivers, especially for poorly performing countries.

The analysis presented here involves the development of mortality rate against motorisation rate over time, spanning a 45-year period, at least for the countries for which all required data are readily available since 1960. The “segmented” package of R software is used, allowing the formation of broken-line regression models (Muggeo, 2005). Such relationships are common in many fields, indicatively including epidemiology and toxicology. In the studied problem this is considered a most appropriate tool, as the motorization rate is empirically identified as an explanatory variable that yields piecewise linear relationships with responses such as traffic risk (i.e. fatalities suffered per motorized vehicle) or mortality rate (i.e. fatalities suffered per head of population).

This work may be viewed as supplementary of other approaches moving to the same direction. For instance, Kopits & Cropper (2004) presented a correlation between the number of fatalities and an area's economic growth. Spline (piecewise linear) functions have been deployed to depict the natural logarithm of mortality rate, traffic risk and motorisation rate over the natural logarithm of real per capita income (GDP). It is worth noting that the per capita income at which traffic fatality risk begins to decline was estimated at \$8600 (in 1985 international prices).

Eksler et al. (2007) have presented a regional analysis of road mortality across Europe. The authors introduced a Bayesian ecological regression model based on the framework of a unified generalised linear mixed model. Population density and country (affiliation) were used as covariates and fitted into the model at the four levels of spatial disaggregation known as “Nomenclature of statistical territorial units (NUTS) regional classification”. A density elasticity of -0.32 was calculated for all countries examined together.

3.2 Data and application

Figure 3.1 presents an overview of the used data for a subset of European countries. This data provides an overview of the available trends and is used as a basis to select the representative countries for the detailed modelling and analysis. A plot of mortality rate (i.e. fatalities per 100.000 inhabitants) against motorization rate is shown, using the same ranges for both axes of all subfigures. This indicator is also known as mortality rate, but the dimension of risk may be more intuitive, since it introduces the meaning of a ratio between (the expected number of) killed road network users and the amount of exposure (in this case: population) –see, for instance, final report of the COST 329 action. In terms of graph interpretation, both terms are considered equivalent, since they yield similar insight. It is important to note that data do not cover the same period of time for all countries, since similar motorization levels were reached at different points in time.

A merely preliminary grouping may be attempted at that point, with several countries exhibiting some turning point. For western European countries this is more pronounced, while for southern ones it is smoother. Eastern European countries seem to evolve in a more time-varying manner, with at least two distinct peaks followed by declining periods of time –since motorisation rate in all cases increases with time.

Figure 3.2 provides a different view of a subset of the same data, this time with time used as the x-axis. The top row of subfigures shows the mortality rate trend for Austria and the UK, both of which exhibit a simple downward trend in the time-span for which data is available. Arguably, if data were available for a longer time period, then the same pattern as that shown in the bottom subfigure for three other EU countries would be apparent.

The remaining subfigures show the mortality rate trend for Belgium, the Netherlands, Greece, Spain and Czech Republic for the period between 1960 and 2005. Several observations can be made already from this figure. First, it appears that the first four of these five countries show a similar pattern of increasing mortality rate until a point, when a structural bend is observed in the trend and a downward trend continues from that point on. The breakpoint in the trend is observed in different points in time. A third (and rather interesting) observation is that, while these breaks seem to be happening at a rather wide range of time (shortly after 1970 for the Netherlands and around 1995 for Greece), they seem to occur at a rather narrow range of mortality rate (somewhere around 25 fatalities per 100.000 population).

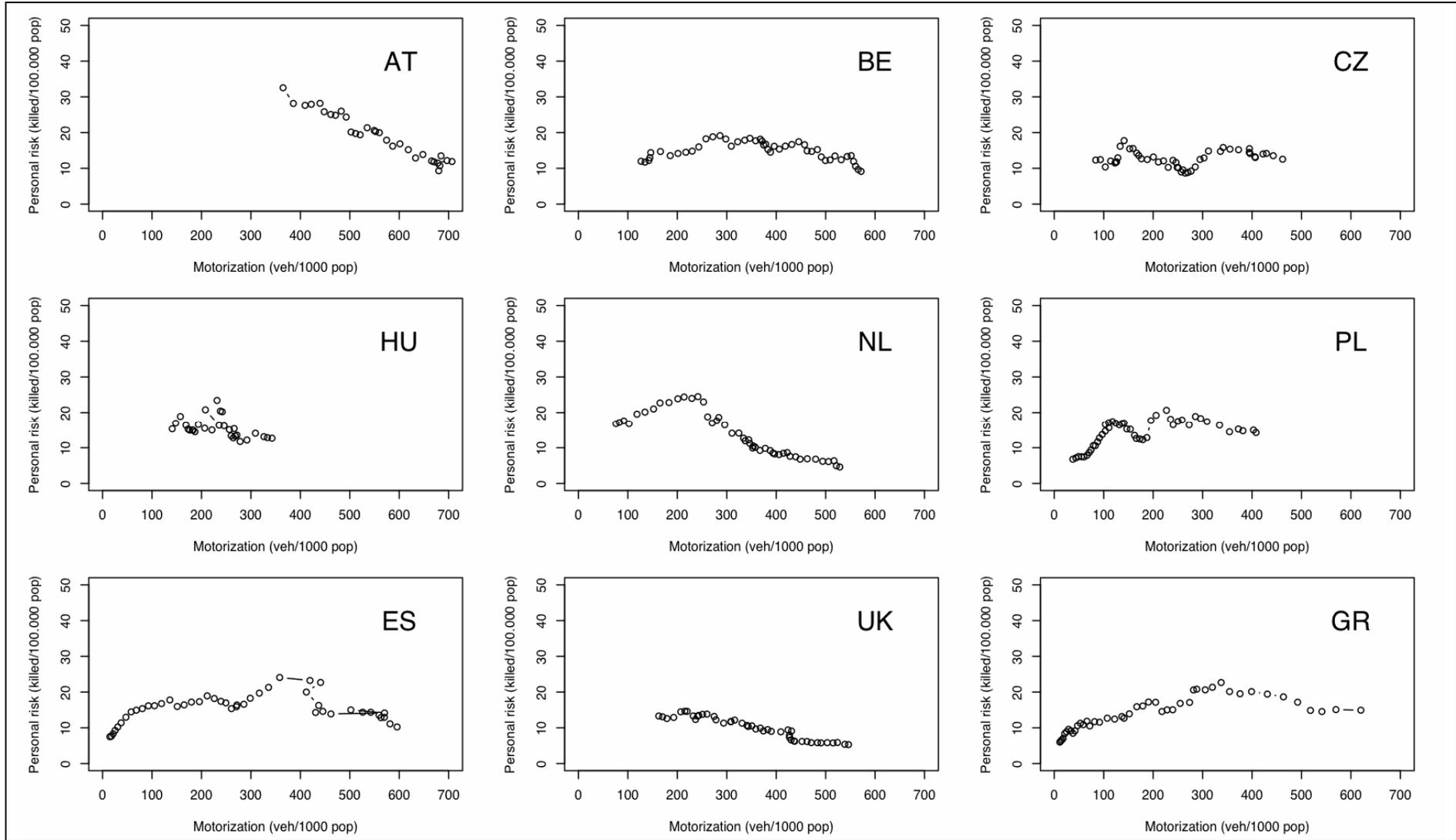


Figure 3.1: Plot of Personal Risk (Mortality rate) over Motorization rate for 9 EU countries (1960-2005)

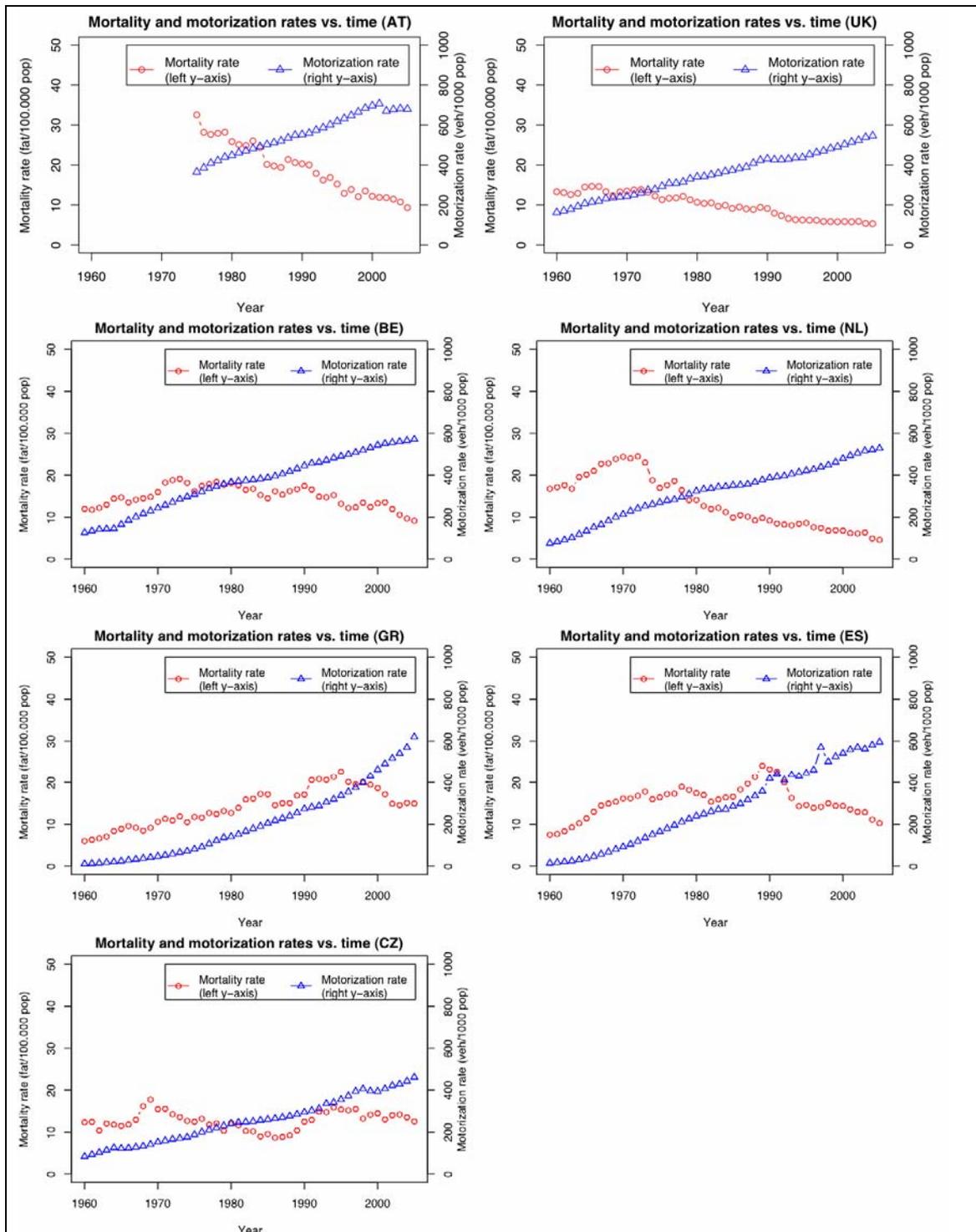


Figure 3.2: Plot of Mortality rate over Motorization rate for selected EU countries (1960-2005). From top left to bottom right: Austria, the United Kingdom, Belgium, the Netherlands, Greece, Spain and Czech Republic.

3.3 Methodology

An excellent synopsis of all relevant analyses that should be combined to the approach presented here may be found in Yannis et al. (2007). The authors make a valuable contribution to the topic by suggesting an evolution of Smeed's Law through autoregressive nonlinear time-series modelling of traffic fatalities.

Data presented in this Chapter involve linear trends and breakpoints. The objective of the analysis is to allow for the simultaneous estimation of regression models with unknown breakpoints in a way that provides estimates both for the breakpoints' locations and the slopes. The advantage of such a methodology is that the main estimates would be consistent, using information from the entire data range and not only from the regime that they span.

The approach that is used is described in detail in Muggeo (2005) and is implemented using the R software for statistical computing with the **segmented** package. The methodology implemented within this package assumes that the number of breakpoints and initial guesses for their values are assumed as input (note that this is very different from assuming that the breakpoints are provided as external inputs!). This is a reasonable and practical requirement that can easily be fulfilled by visual inspection of the data.

While a detailed statistical presentation of the used method is not within the scope of this section, a brief description of estimation and testing processes for models developed with R's segmented package is provided, following the exposition used in Muggeo (2005). Muggeo (2003, 2005) can also be used by the interested reader as sources for further details.

A segmented relationship between the mean response $\mu = E[Y]$ and the variable Z , for observation $i = 1, 2, \dots, n$ is modelled by adding in the linear predictor the following terms:

$$\beta_1 z_i + \beta_2 (z_i - \psi)_+ \quad (3.1)$$

where $(z_i - \psi)_+ = (z_i - \psi) \times I(z_i > \psi)$ and $I(\cdot)$ is the indicator function equal to one when the statement is true. According to such parameterization, β_1 is the left slope, β_2 is the difference-in-slopes and ψ is the breakpoint. We tacitly assume a GLM with a known link function and possible additional covariates, x_i , with linear parameters δ , namely $link(\mu_i) = x_i' \delta + \beta_1 z_i + \beta_2 (z_i - \psi)_+$; however, since the discussed methods only depend on (3.1), we leave out from our presentation the response, the link function, and the possible linear covariates.

The package **segmented** offers facilities to estimate and summarize generalized linear models with segmented relationships; virtually, no limit on the number of segmented variables and on the number of change-point exists. **Segmented** uses a method that allows to estimate simultaneously all the model parameters yielding also, at the possible convergence, the approximate full covariance matrix.

With respect to testing for a breakpoint, if the breakpoint does not exist, then the difference-in-slopes parameter has to be zero. In this case, a natural test for the existence of ψ is

$$H_0 : \beta_2(\psi) = 0 \quad (3.2)$$

Note that here we write $\beta_2(\psi)$ to stress that the parameter of interest, β_2 , depends on a nuisance parameter, ψ , which vanishes under H. It is easy to guess that several non-regularities feature such a hypothesis and, as a consequence, standard statistical tests (e.g. Wald) may not be applied.

3.4 Estimated models

This section starts with a presentation of the “simpler” trends, observed e.g. in Belgium and Greece, before exploring examples of gradually more involved relationships. Each figure contains three subfigures. The top subfigure presents the mortality rate data versus the motorization rate, and the estimated model. The next subfigure shows the same data and estimated model but uses a different x-axis for projection, i.e. the time. Even though the fitted model may appear to be non-linear in this subfigure, it is clarified that this is simply a result of the mapping of the explanatory variable (motorization) into the time axis. Finally, the bottom subfigure presents motorization rate, i.e. the evolution of the vehicle fleet at national level over time.

Belgium (Figure 3.3) and Greece (Figure 3.4) show the simpler pattern of an increasing trend until a maximum/breaking point is reached and a downward trend starts. In the case of Belgium the break point is recorded by the time the motorization rate reached the value of 320 vehicles per 1000 inhabitants (in 1976). Interestingly, despite its definitely improving trend, Belgium only reached the level of mortality rate experienced in early 1960's by the year 2003; that is, after 43 years. Of course, it should be stressed that the motorization rate at national level in 1960 was already 127 vehicles per 1000 inhabitants.

See Appendix A3.1 for the Tables presenting parameter coefficients for all countries (one Table per country, including break points estimates, followed by coefficients for intercepts and slopes).

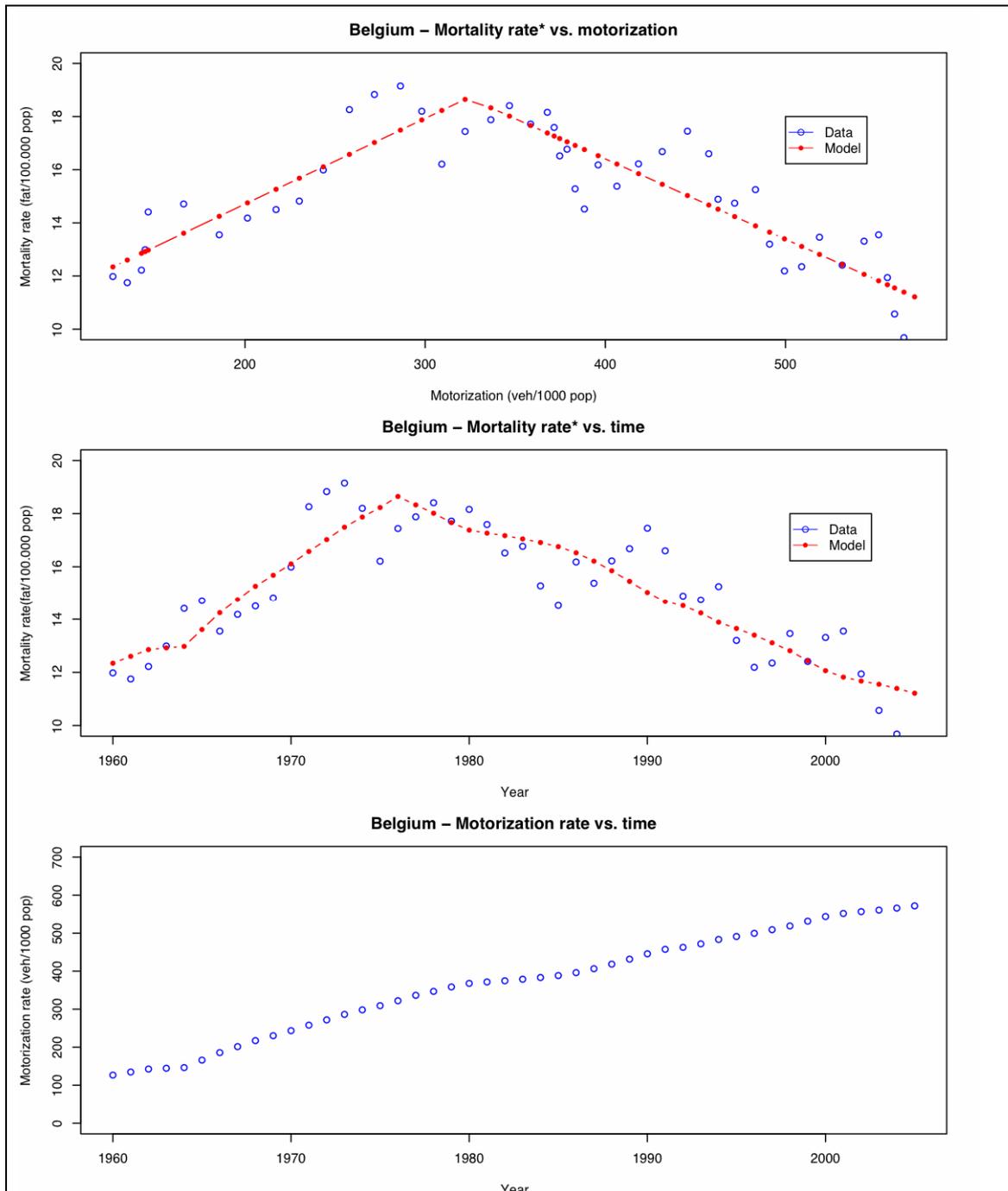


Figure 3.3: Estimated model for Belgium (*using “killed on the spot” definition). Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

Figure 3.4 presents the same information for Greece. The structural break-point is observed at a motorization rate of 335 vehicles per 1000 inhabitants. While this number is not much different than the respective number of Belgium, it occurred almost 20 years later.

It is noted that mopeds are not included in the motorization rate (to maintain comparability with the data from other countries). The actual values may in fact be slightly different if one includes mopeds in the total vehicle fleet; an interesting choice given that mopeds appear to participate in serious injury and

fatal accidents to a non-negligible extent, especially when other European countries are also considered. This is indeed a subtle point that calls for attention in cases where some generally unusual mode becomes rather common. Netherlands could also serve as an example, for the case of bicycles.

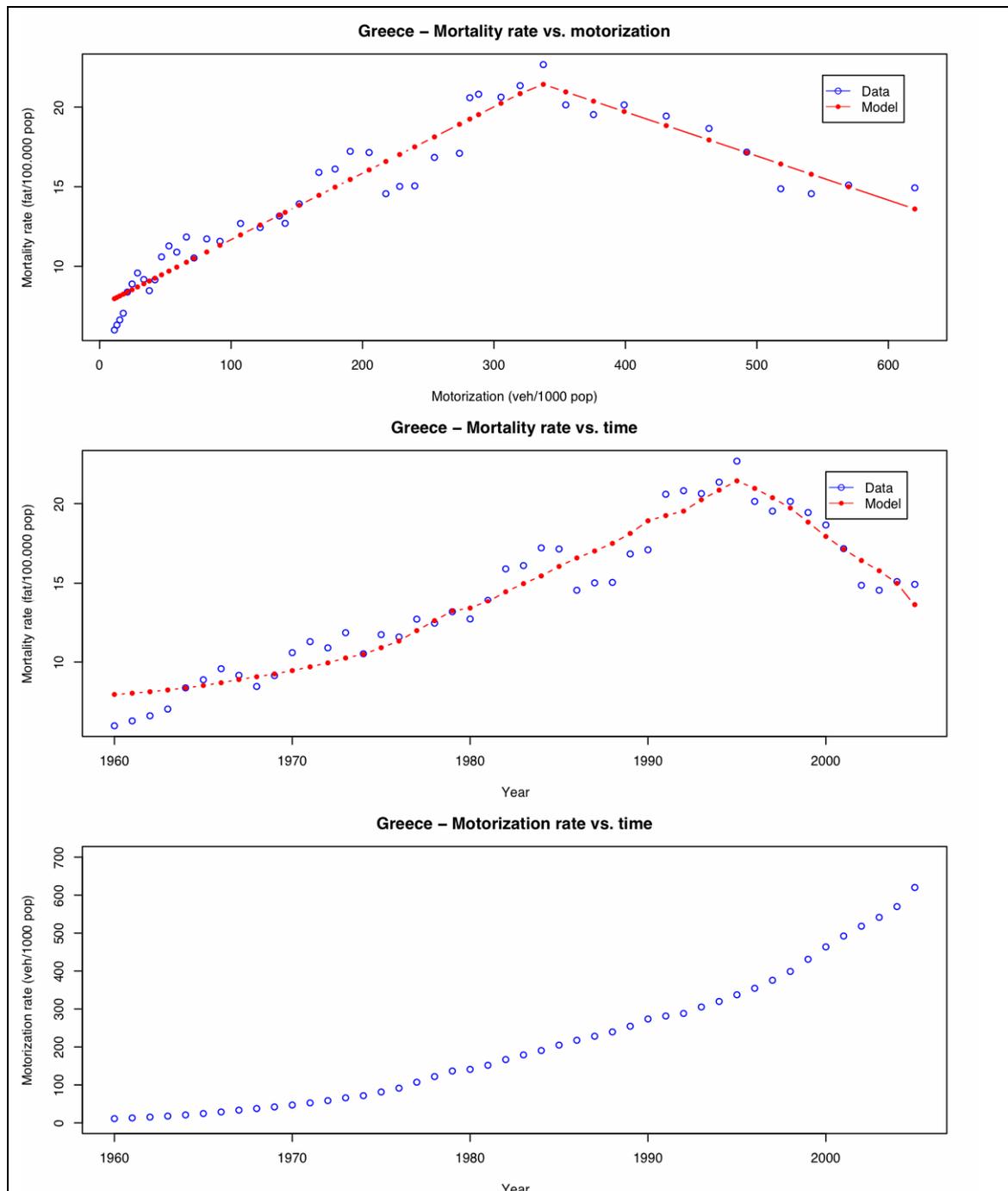


Figure 3.4: Estimated model for Greece. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

In the case of Netherlands the break point is recorded even earlier than Belgium, around 1972. More impressively, this year corresponds to a value of motorization rate around 185 - a relatively low value compared to most other cases. Although a neighbouring country with Belgium, Netherlands only needed

18 years to return to the mortality rate suffered in 1960 (in the order of 16.5 fatalities per 100,000 people in the year 1978). The motorization rate in 1960 was around 76 vehicles per 1000 inhabitants –much lower than in neighbouring Belgium, for instance, by that time.

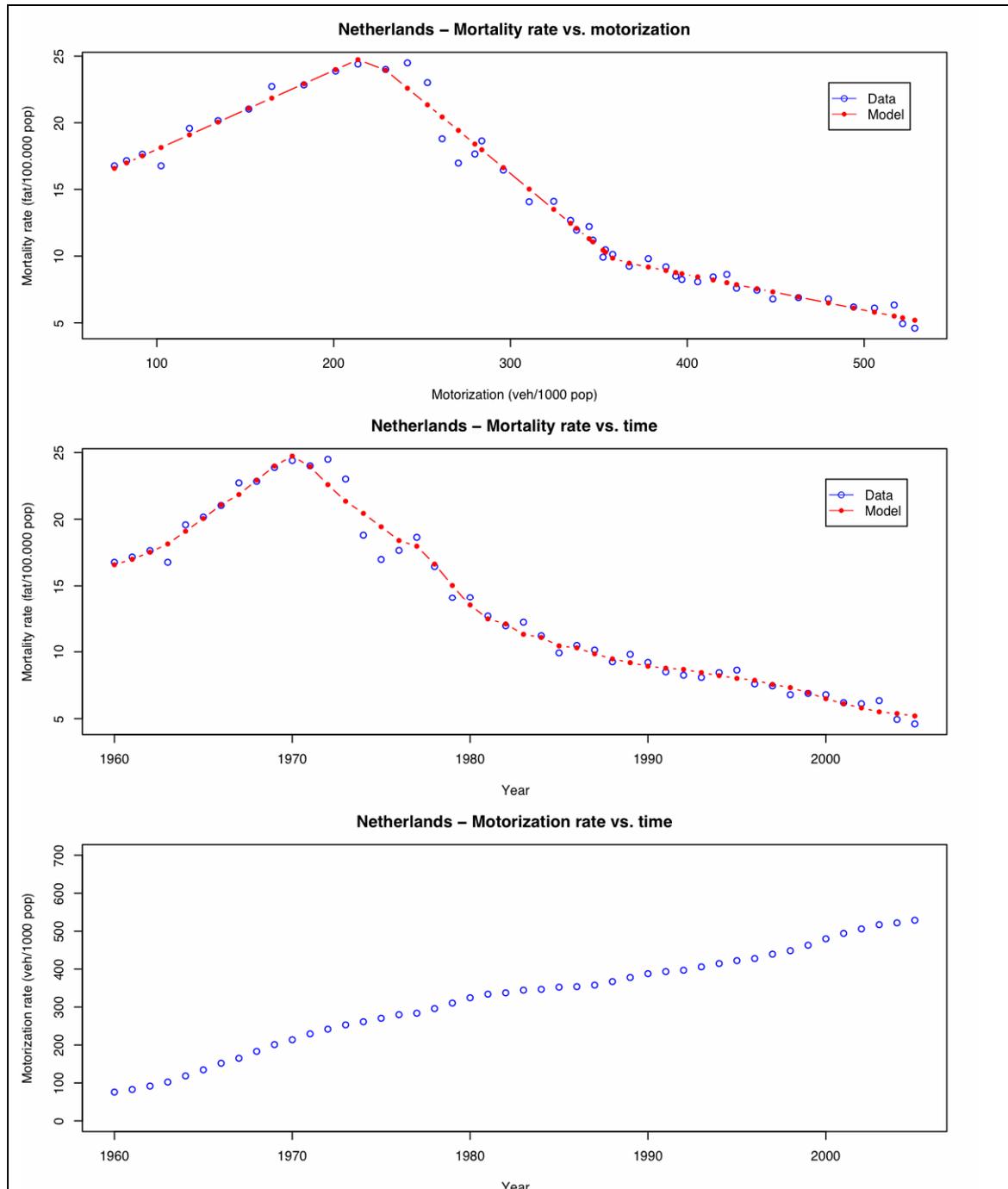


Figure 3.5: Estimated model for the Netherlands. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

The data from the Netherlands indicate that a secondary breakpoint took place after the mortality rate was reduced to about 10 fatalities per 100.000 inhabitants. The rate of decrease of mortality rate changed (became lower) after that point. This is an interesting and intuitive finding: as mortality rate gets

lower, there is less “room” for improvement. As stated by several researchers from Smeed’s time up to present (e.g. Lassarre, Page etc), each basic road safety-related measure reasonably contributes to notable improvement in the early stages of its introduction, fading out as the system reaches a new equilibrium.

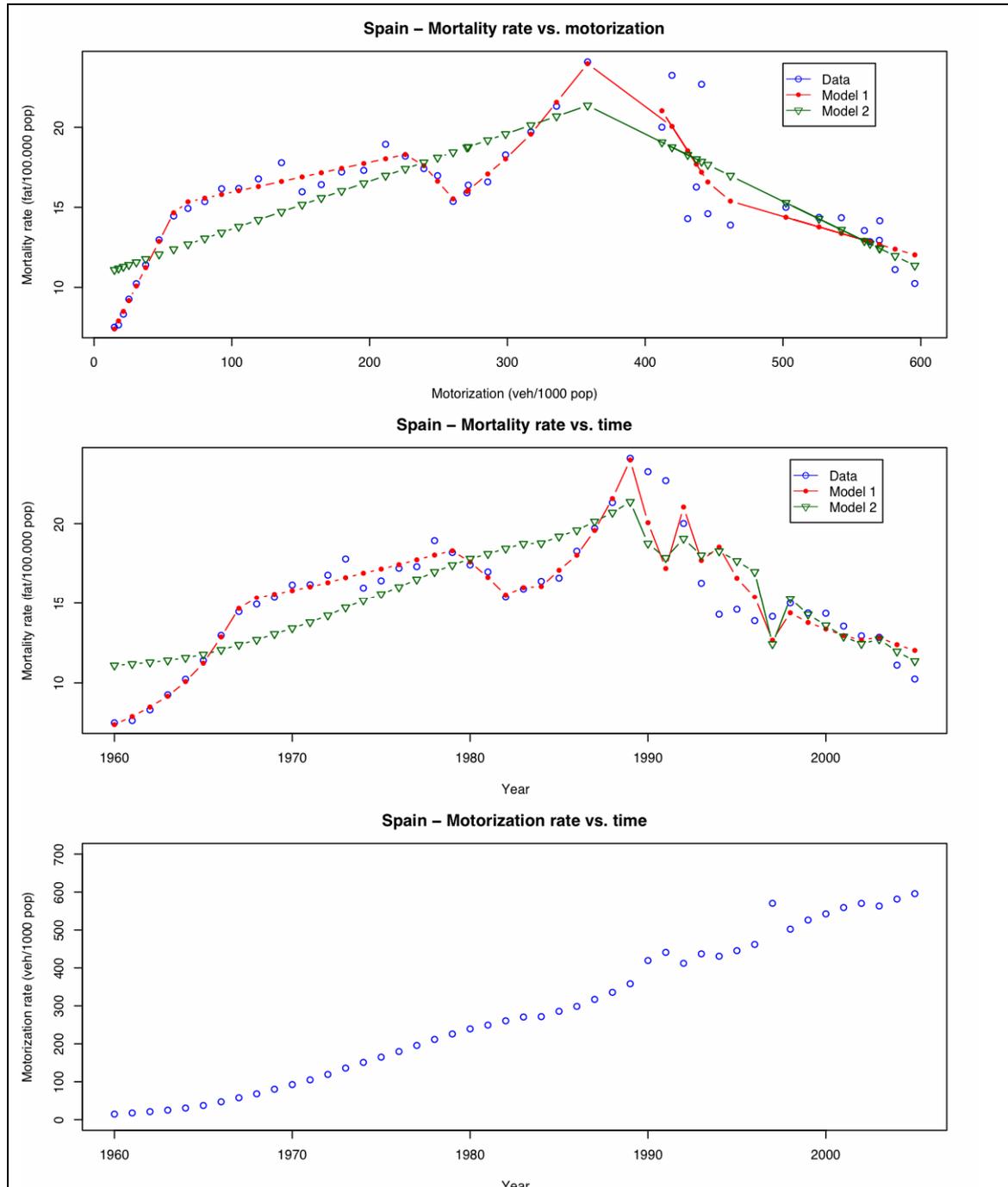


Figure 3.6: Estimated model for Spain. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

Figure 3.6 presents the estimated models for Spain. After careful examination of the data points, two alternative models have been developed for Spain: one parsimonious (revealing one single break point), and one involving several breakpoints. The former, simpler model captures the overall trends and

breakpoint, and is consistent with the models developed for the previous countries; however, it may miss some subtle patterns and local phenomena. The latter may be more accurate in mathematical terms, but it is much more complicated to interpret. Therefore, it has to be examined in a more complete manner, presumably in parallel with the history of major road safety-related programmes or specific measures.

If secondary effects (captured by the detailed model) are put aside at the moment, one may observe that the main response of Spain to the increased mortality associated to the increase of vehicle fleet came somewhat late, i.e. when motorization rate exceeded the threshold of 370 vehicles / 1000 inhabitants. Again, in order to make a meaningful comparison it is necessary to examine the composition of the vehicle fleet –whether it includes the same classes as in the other countries.

Czech Republic reveals a more complex trend, with two clear consecutive peaks and a total of three breakpoints. Mortality rate is increasing from very low motorization levels and the first break-point appears for a motorization rate of about 150 vehicles per 1000 inhabitants. After that point mortality rate decreases linearly with motorization, until a threshold of 275 vehicles per 1000 inhabitants, when mortality rate started to increase again. After another critical value of 320 vehicles per 1000 inhabitants, mortality rate started to decrease again. This complex image could be related to the change that occurred in the political regime at some point. Mortality rate at country level started picking up again in the second half of the 80s, even more rapidly than in early years.

Earlier in this section, it was hypothesized that there is a critical value of mortality rate, which results in a breakpoint in its evolution. This could, e.g. be related to the magnitude of the problem, resulting in a widespread mobilization and sensitivity about the problem of road safety, leading to some effective measures. The two consequent peaks/break-points that occurred in the Czech Republic support this hypothesis as they happened at essentially the same level of mortality rate. Values across countries vary due to many other unobserved variables and parameters, but this limited empirical observation is interesting.

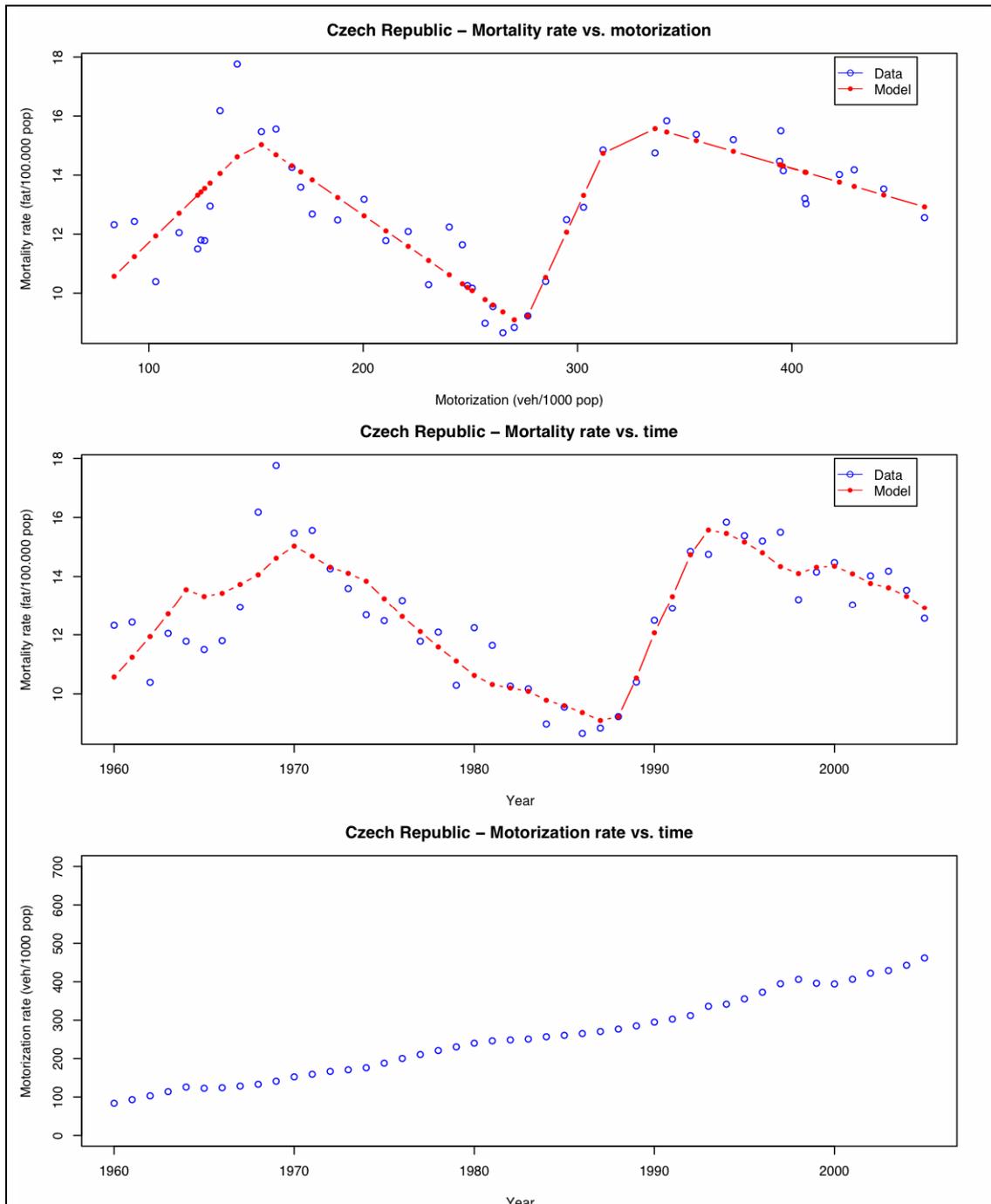


Figure 3.7: Estimated model for the Czech Republic. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

3.5 Conclusion

In the previous sections of this Chapter, the trends of road traffic fatalities in several EU countries have been examined given the evolvement of elementary socioeconomic indicators, namely motorised vehicle fleet and population, at country level. Piece-wise linear functions are selected for this analysis by means of the segmented package of the R software, as allowing for flexibility while treating data at aggregate (national) level. The motorization rate, i.e. the

number of vehicles per 1,000 inhabitants, has been selected as independent variable, since it incorporates both parameters of interest.

It is possible to view graphs concerning all countries together in one figure, i.e. Figure 3.8. This is important, since it depicts in a comprehensive manner that:

- Different countries reached specific motorization rates at distant moments in time (temporal landmarks)
- Some of those countries exhibit a break point within a narrow range of motorization rate values, implying perhaps similar social and economic conditions
- This range is different for certain subgroups among the examined countries, providing a hint that some grouping may be of meaning in geographic and socioeconomic context

Possible interpretation of the obtained results is treated with some reserve, as motorization rate may be defined in different ways. The nominator of this figure (fatalities) may be regarded more or less well-defined, after many efforts put at pan-European level for a common definition (30-day fatalities). As far as the denominator is concerned, available data of vehicle fleet shows some slight discrepancies, e.g. the total number of vehicles in Spain reveals some irregular steps for specific years. Furthermore, each vehicle class is ruled by specific particularities, presumably implying a camouflage for systematic errors (Katsochis et al., 2006). For instance, the unusually large number of registered mopeds in Greece is not necessarily equivalent to circulating ones; and it does not constitute a simple issue to ascertain recording of abandoned or in any way not driven vehicles at national level. The application of common definitions should be further examined, so that there is an as common as possible base for comparison.

This approach involves mortality rate, a seemingly simplistic indicator according to several researchers (e.g. Page, 1997). On the other hand, it seems practically necessary to return to the examination of such trends, so as to obtain more complete insight and background prior to the analysis of precise exposure data (vehicle-km driven per vehicle type), when credibly available. An initial classification of European countries is attempted at this stage.

This effort may and should be followed by similar groupings involving specific vehicle types and population subsets (age groups or gender). It will then be much easier to distinguish cases and consider the presence of true impact due to GDP or other growth-related parameters; so, it is not advised to neglect the study of elementary indicators, especially when difficulties are encountered in the reliability of more exposure-data oriented analyses.

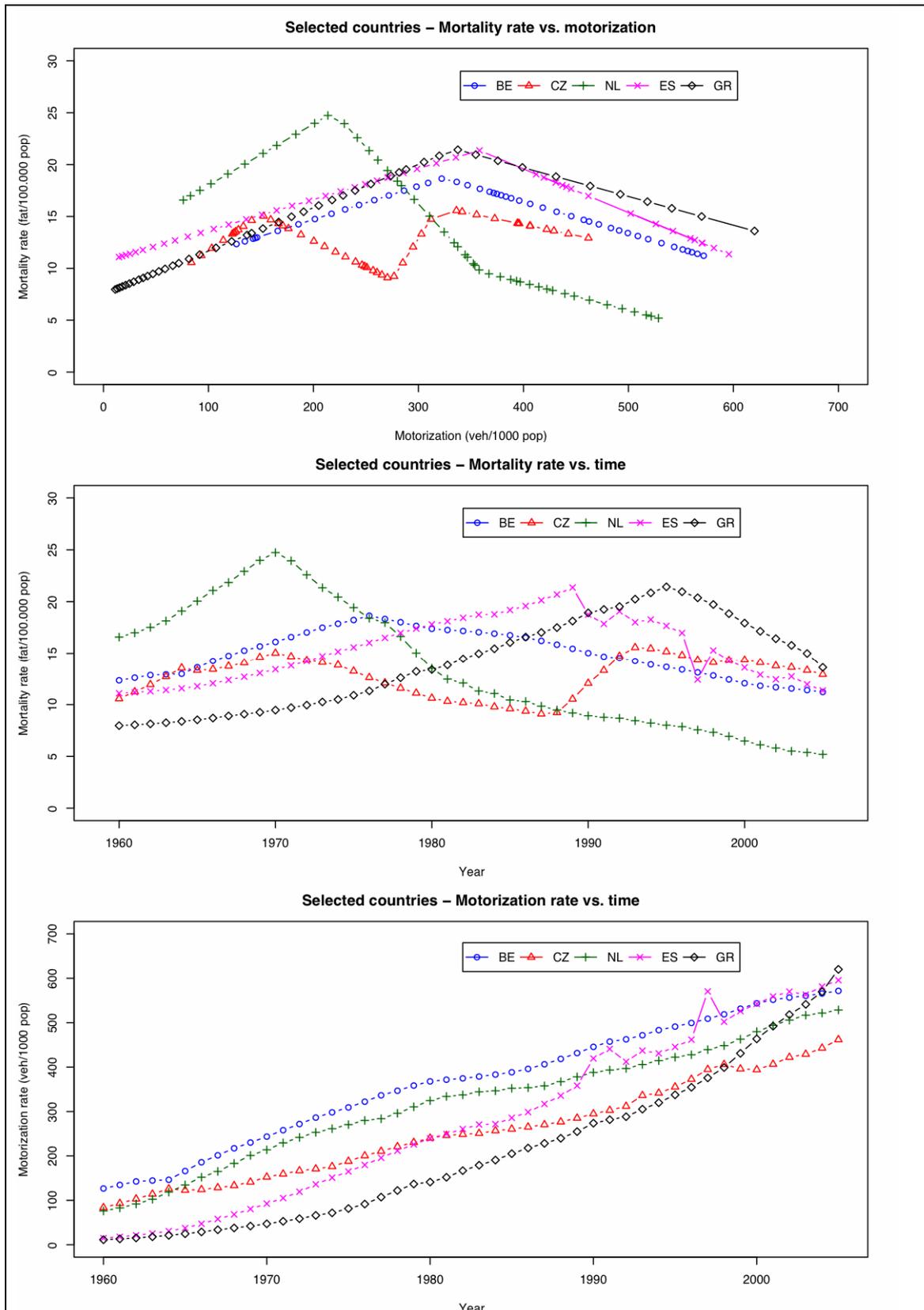


Figure 3.8: All estimated models. Three subfigures: Top: mortality rate vs. motorization, in the middle: mortality rate vs. time, bottom: motorization vs. time.

Furthermore, such a course of work will allow facilitating to a larger extent the results obtained from other analyses. For instance, as already mentioned in the introductory part of this Chapter, Eksler et al. (2007) studied data from 25 EU member states and concluded that, on average, a 10% increase in population density is linked to a 3.2% decrease in road fatalities. It would be interesting to have an estimate of this effect's part to be appointed to vehicle fleet-related trends, as densely populated areas often tend to develop in a varying manner.

Based on the profile of substantially different countries, it is reasonable for one to assume that the trend in question is practically universal in the EU –solely as a scheme, at the moment. More explicitly, in all cases there is some structural relation of the suffered fatalities per population (mortality rate) with the motorization rate expressed as number of motor vehicles per population.

The trend in question does not occur simultaneously in all examined countries. This is expected, since a given motorization rate is reached at a different time horizon for each country. In addition to that, the value of motorization rate for which the break point is observed also varies across countries, further complicating this phenomenon. It is considered critical to identify why this is so, mainly by comprehending what the lag between two countries captures or represents. This will facilitate the comparison of specific pairs of countries.

It is noted that the explanatory variable is motorization rate rather than time. However, both for interpretation and presentation purposes, it makes sense to show a secondary x-axis referring to time. This will be included in a later version of the document. Obviously, only one of the two axes will be equi-distant, so it first needs to be decided if this will be motorization or time.

It is also stressed that the output of the models developed in this work includes values and Standard errors for breakpoints and slopes, so all the information basically required to draw a first comparison is available. The results obtained so far may be perceived as a proof of concept that captures the structural changes etc. Still, if the objective of the research is to be fulfilled, it is necessary to find a way to check/clean-up residuals, which appear correlated at the time.

These preliminary findings should serve as an adequate base to pursuit further understanding of why and when such break points are observed. Research already conducted in this field will be taken into account so as to facilitate a useful grouping of the examined countries. The ultimate objective of next steps in the framework of this research should be to utilize this finding in order to make reliable predictions for countries or territories for which the so-called “break” has not occurred yet.

3.6 References

Al-Haji, G. (2007). Road Safety Development Index (RSDI). Theory, Philosophy and Practice. Linköping Studies in Science and Technology, Dissertation No. 1100, Norrköping, Sweden.

COST 329 (2004). Models for traffic and safety development and interventions. Final Report of the Action. European Commission, Luxembourg.

Eksler, V., Lassarre, S. & Thomas, I. (2007). Regional analysis of road mortality across Europe. *Journal of the Royal Institute of Public Health*, Article in Press (accepted 16 October 2007).

Hinkley, D.V. (1971). Inference in Two-Phase Regression, *Journal of the American Statistical Association*, Vol. 66, No. 336, pp. 736-743

Katsochis, D., Handanos, Y. & Yannis, G.(2007). Comparative assessment of road safety performance in Greece. *Proceedings of the 11th World Conference on Transport Research (WCTR)*. Berkeley, CA, USA – June 2007.

Kopits, E., & Cropper, M. (2005). Traffic fatalities and economic growth. *Accident Analysis and Prevention*, Vol. 37, pp. 169-178.

Lassarre, S. (2001). Analysis of progress in road safety in ten European countries. *Accident Analysis and Prevention*, Vol. 30, pp. 743-751.

Muggeo, V. (2005) segmented: an R software to fit regression models with broken-line relationships. Dipartimento Scienze Statistiche e Matematiche 'Vianelli' Università di Palermo, Italy. vmuggeo@dssm.unipa.it

Muggeo, V. (2003) Estimating regression models with unknown break-points. *Statistics in Medicine*, 22: p.p. 3055-3071.

Smeed, R.J., 1968. Variations in the pattern of accident rates in different countries and their causes. *Traffic Engineering Control*, Vol. 10(7), 364-371.

Twisk, D.A.M. & Stacey, C. (2007). Trends in young driver risk and countermeasures in European countries. *Journal of Safety Research*, Vol. 38 (2007), p.p.245–257.

Yannis, G., Antoniou, C. & Papadimitriou, N. (2007). Autoregressive nonlinear time-series modeling of traffic fatalities in Europe. Article under review.

Chapter 4 - Disaggregation by age and gender

Jacques J.F. Commandeur, Frits D. Bijleveld, Henk L. Stipdonk (SWOV)

4.1 Introduction

In Chapter 2 we showed how the development of traffic safety can be described by the development of mobility and the development of risk: the number of accidents or fatalities per distance travelled. Changes in risk can be due to changes in circumstances or safety measures, but can also be caused by a shift in the use of traffic modes, road types, etc. The macroscopic development of risk can e.g. be influenced by a shift from unsafe traffic modes to safe traffic modes or from unsafe roads to safe roads. Also, when the fraction of female drivers increases, or the fraction of unexperienced drivers decreases, this can also cause a change in the macroscopic risk.

One way to obtain a better understanding of the global macroscopic development in a country's road safety is to investigate the development of subgroups of the road user population. In this chapter we discuss (future) developments in road safety while focussing on two demographic aspects of the road user population: their age and their gender. It is well-known that young inexperienced car drivers are more at risk of being involved in traffic accidents than older more experienced car drivers; in general male drivers are also more at risk than female drivers (see for example Van den Bossche (2006), and Van den Bossche, Wets, and Brijs (2007)). If the distance travelled by young (male) car drivers decreases over time, for example, then this will affect the number of casualties not only in that age group but in the other age groups as well, since their chances of being involved in an accident involving young (male) drivers will become smaller. All this implies that when the composition of a country's population changes over time this will affect the composition of that country's mobility on roads and may therefore also affect the number of accidents and their outcomes.

For an analysis of the effect of distance travelled by age and gender, mobility data disaggregated by age and gender have to be available. For The Netherlands, we have such data at our disposal since 1985. Until 2004 these data are available for age groups, and from 2004 onward they are available for individual age years. Of course, the last is to be preferred. When data for age groups are available only, a potential problem is that possible *cohort effects* in the development of the individual mobility of age groups cannot be recognized easily. Cohort effects could occur when e.g. the average distance travelled by road users of 30 years old in 1985, say, was different from the average distance travelled by road users of 30 years old in 2001, for example. A similar effect may also exist for risks.

In this Chapter we investigate the possibility and usefulness of incorporating data on population size for the analysis of developments in road safety

disaggregated by age and gender. In the following section we first discuss how information on the population size can be used to break down the mobility figures for age classes into age years. Also, and even more importantly, we show how the information on mobility figures and population size can be combined to obtain forecasts for the mobility of road users disaggregated by age class and -if required- by age year. Then, in the next section the models and some selected highlights of the results of time series analyses exploiting these results are presented, primarily showing the differences between forecasts of accident outcomes excluding and including data on population size disaggregated by age and gender. In the last section, we give the main conclusions.

4.2 Decomposition of mobility into population size and mobility per capita

In this section we show that population figures can be used to get a better understanding of the mobility figures in the past, also disaggregated per age class or year. To check the effectivity of this procedure, we compare in -ample forecasts with the data available. Forecasts, based on the data from 1987 to 2000 are used to predict the mobility in 2007, and compared with the data.

In the Netherlands, an estimate of the annual distance travelled by road users is obtained by means of an annual survey called MON (Mobiliteitsonderzoek Nederland, in English: Mobility Research Netherlands). Until recently, the data on distance travelled in this survey was published for the following 13 age *groups* or *classes*: 0-5 years, 6-11 years, 12-14 years, 15-17 years, 18-19 years, 20-24 years, 25-29 years, 30-39 years, 40-49 years, 50-59 years, 60-64 years, 65-74 years, and 75-99 years. At the same time, Statistics Netherlands (CBS) provides annual figures for the total number of inhabitants of the Netherlands including disaggregated numbers for each age *year* (not age class) of the Dutch population.

The annual population size per age class is a fairly good predictor of the annual distance travelled by all road users per age group as obtained from the MON survey. This can be concluded from performing a regression analyses for the total distance travelled y_{ct} (all traffic modes together) in year t in each of the just mentioned 13 age classes c , where y_{ct} is a linear function of t , and proportional to x_{ct} , the annual population size for age class c (as obtained from the CBS):

$$y_{ct} = x_{ct}(a_c + b_c t) + e_{ct} \quad (4.1)$$

for $t = 1, \dots, n$. In (4.1), a_c and b_c are unknown parameters, and e_{ct} is the residual for age class c in year t . It follows from (4.1) that the regression line $(a_c + b_c t)$ actually represents the estimated distance travelled by all road users *per capita* in age class c . A weighted version of regression model (4.1) was applied to each of the 13 age classes, the weights consisting of the accuracy of the survey figures as expressed by their standard deviations for each age class. The results of these analyses are presented in Appendix A4.1. Figure 4.1 and Figure 4.2, show an example of just 2 of these 13 separate age groups.



Figure 4.1: Annual distance travelled by all male or female inhabitants of age group 18-19, both data and model. Left hand panel: total distance travelled (10^9 km, dots) and 95% confidence intervals for the estimate. The thick line represents the prediction by model (4.1). Right hand panel: per capita distance travelled (10^9 km) per million inhabitants of age group 18-19 according to the model.

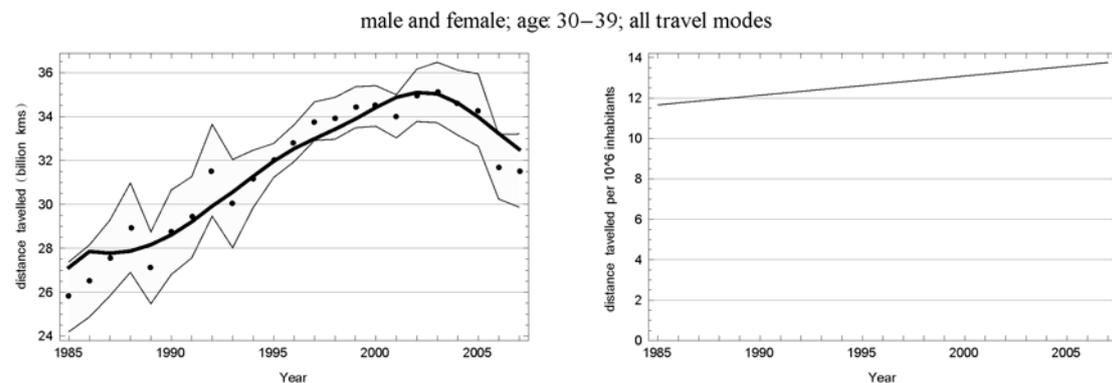


Figure 4.2: Annual distance travelled by all male or female inhabitants of age group 30-39, both data and model. Left hand panel: total distance travelled (10^9 km, dots) and 95% confidence intervals for the estimate. The thick line represents the prediction by model (4.2). Right hand panel: per capita distance travelled (10^9 km) per million inhabitants of age group 30-39, according to the model.

The left panels of Figure 4.1 and Figure 4.2 show the observed annual distance travelled by all road users for each age class (indicated by dots), together with their 95% confidence intervals (as obtained from the standard deviations provided by the MON survey), and the predictions of the annual distance travelled by all road users for each age class (indicated by thick a continuous line), based on the model. The right panels of Figure 4.1 and Figure 4.2 display the estimated annual distance travelled by all road users *per capita* for each age class, as expressed by $(a_c + b_c t)$ in (4.1).

For the group of 18-19 year old the per capita annual distance travelled is almost constant in the observed period. The total annual distance travelled was decreasing due to a decreased population size.

For the group of 30-39 year old, the per capita annual distance travelled was increasing in the observed period. However, the total annual distance travelled increased at first but decreased after 2002, due to changing population size.

Figures 4.1, 4.2 (and Figure A4.1) show that for each age class, the observed annual distance travelled by all road users is described quite well from the product of the actual annual population size and the estimated annual distance travelled Model (4.1) was also used to analyse further disaggregations of the traffic process (i.e., also to several types of combinations of age, gender and road user type). The results are presented in Appendix A4.1. These figures confirm the general idea that the decomposition of the annual distance travelled in an age class into the annual population size of that age class and the estimated distance travelled per capita of that age class works quite well for further disaggregations also.

The MON survey not only provides estimates of the total annual distance travelled per age class but also for different modes of transport. For example, if we only consider mobility by car drivers, then the annual distance travelled by car drivers for all 13 age classes can be combined to obtain the three-dimensional plot shown in Figure 4.3.

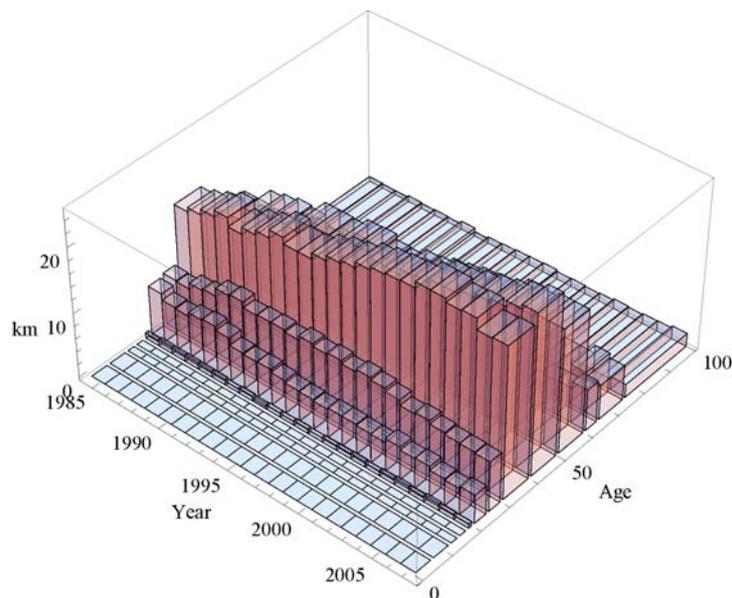


Figure 4.3. Annual distance travelled (in 10^9 km) by car drivers by age group (thirteen classes) and calendar year (1985-2007).

This figure actually contains 23 histograms (one for each calendar year 1985, 1986, etc.) consisting of 13 age classes stacked one beside the other on the time axis. The distance travelled by the four age classes 0-5 years, 6-11 years, 12-14 years, and 15-17 year are all zero, for the obvious reason that persons younger than 18 years are not allowed to drive a car in The Netherlands.

The annual mobility data of car drivers available for all age classes in Figure 4.3 and the annual data on population size of all age years (as available from Statistics Netherlands) were combined to obtain smooth estimates of the annual distance travelled by car drivers in 1985-2007 for each age year (and therefore

for car drivers of 18 years old, of 19 years old, of 20 years old, etc.) rather than for age groups.

This was achieved by optimising the agreement between the model and the data by means of minimising the so-called loss function:

$$f(a_{ij}, b_{ij}) = \sum_{t=1}^n \sum_{j=1}^C w_{jt} \left(y_{jt} - \sum_{i=1}^{c_j} x_{ijt} (a_{ij} + b_{ij}t) \right)^2 + \text{penalty}, \quad (4.2)$$

with respect to all the intercepts a_{ij} and regression coefficients b_{ij} , where C is the number of age classes, and c_j denotes the number of age years in age class j ($j = 1, \dots, C$). Weights w_{jt} are used to account for differences in accuracy of the data. The practical effect of including the penalty function in the loss function (4.2) is that it is minimized in such a way that the estimated distance travelled per capita will be similar both for consecutive time years *and* for consecutive age years. Thus, the quite reasonable assumption is made that these developments over consecutive time years and consecutive age years are relatively smooth. Details about the penalty function are found in Appendix A4.2.

In order to optimise model (4.2), information of mobility data per individual age year available for the years 2004-2007 are used. We used mobility data per age category up to 2003 to predict mobility data per individual age year for 2004 and 2005. Smoothing parameters were chosen such that model (4.2) optimally predicts mobility data per individual age year for 2004 and 2005. The results were verified using mobility data per individual age year for 2006 and 2007 which were not used in that analysis. The results for 2004 and 2005 were very similar to those for 2006 and 2007. The resulting smoothed estimate of individual mobility as car driver per age year is given in Figure 4.4.

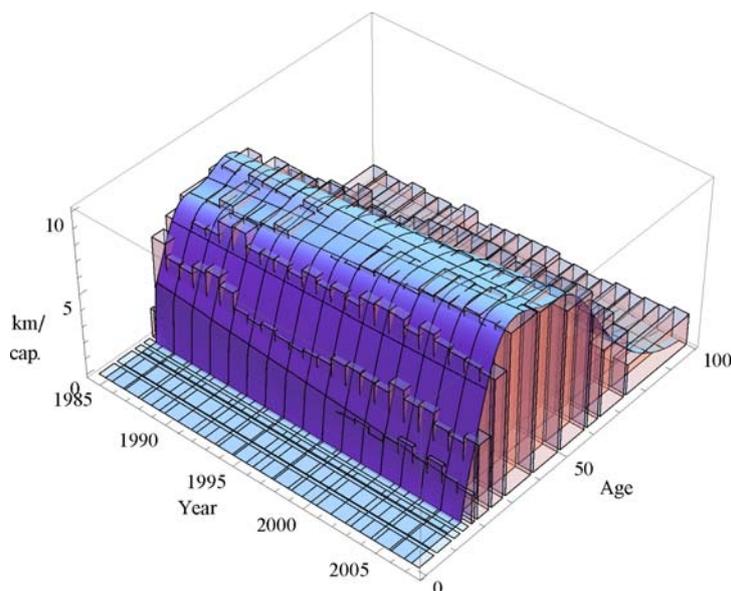


Figure 4.4. Smooth estimates (blue plane) and original data (red histogram) of distance travelled per age year for the years 1985-2007.

Then, using the coefficients obtained from the analysis using mobility data per age category up to 2003, model (4.2) was used on mobility data per age category up to 1993 only. The results were used to predict mobility per individual age year for the years 2004-2007. The results of this analysis are given in Figure 4.5.

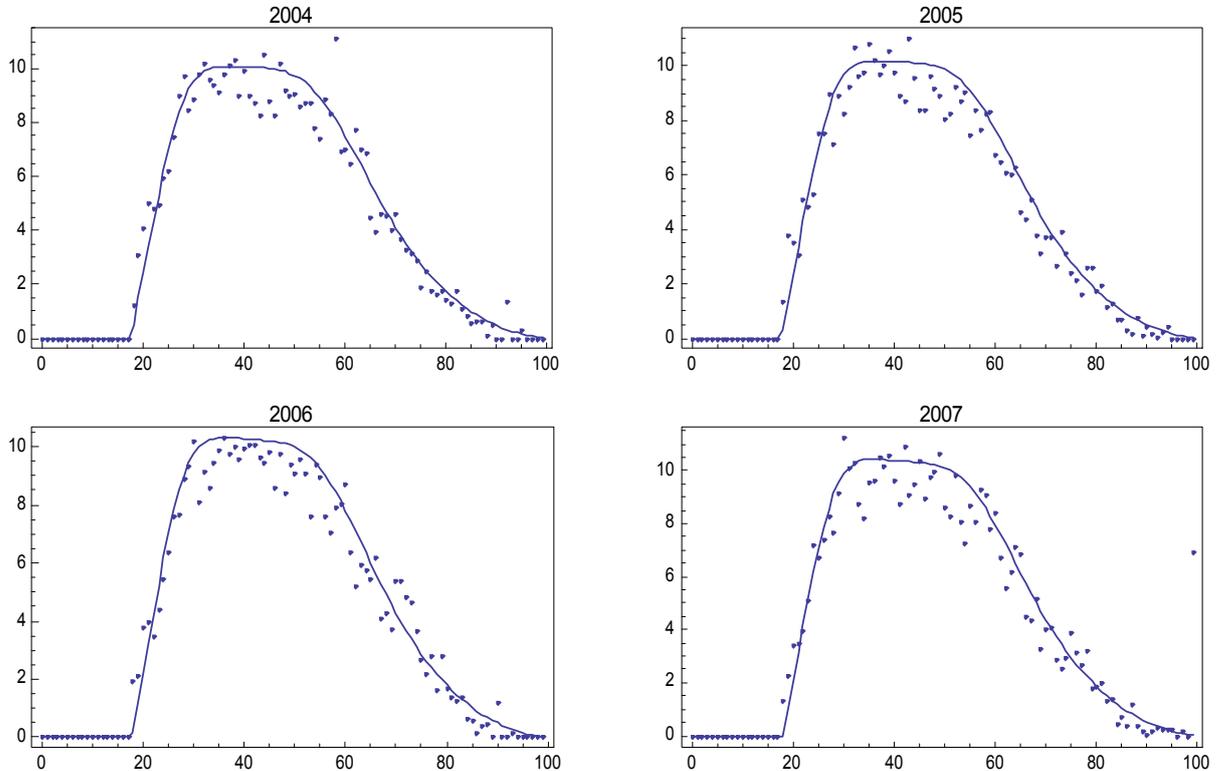


Figure 4.5. Forecasts (line) and original data (dots) of distance travelled by car drivers per age year in 2004 through 2007. The forecasts are based on the analysis of the time period 1985-1993..

Figure 4.5 shows that the mobility figures per age year of the Dutch MON survey in 2004-2007 enables a straightforward comparison of the forecasts with the actual mobility figures. Figure 4.5 indicates that the model recovers the actual mobility figures more than ten years after the time frame used in the analysis, even though the mobility figures in 1985-1993 are only available for age classes of car drivers. This suggests two important characteristics of mobility:

- The per capita car driver mobility is a smooth function of age
- The per capita car driver mobility is a slowly and predictably changing function of time.

The analysis has shown that the use of population data simplifies a good understanding of mobility by age. The availability of population data enables us to improve the forecasts of mobility. In The Netherlands, Statistics Netherlands issues actual forecasts for the development of the Dutch population per age year. These forecasts – which extend further than the year 2020 – together with the parameter estimates of model (4.2) applied to the period 1985-2007 were

used to predict the distance travelled by car drivers per age year, the results of which are presented in Figure 4.6.

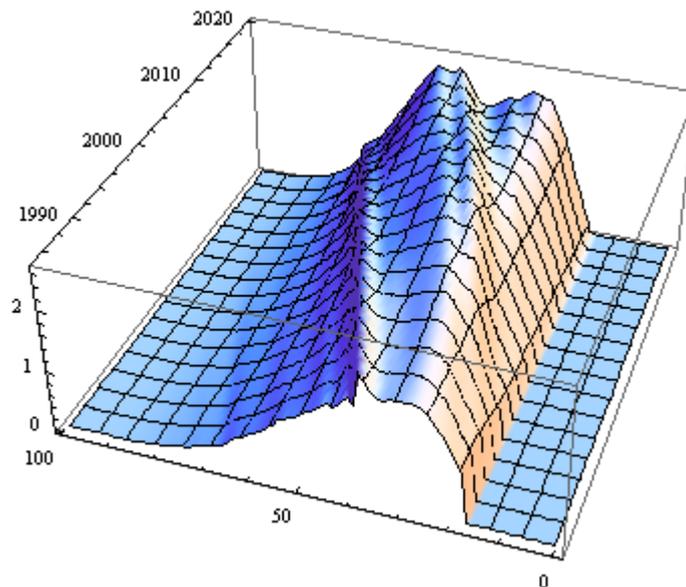


Figure 4.6. Estimated annual distance travelled (10^9 km) by car drivers against their age year (from 0 to 100 years old, from right to left) and time (front to back), including forecasts until the year 2020.

Figure 4.6 displays some interesting features.

- The development of distance travelled by car drivers over time contains an upward 'wave' on the left part of the Figure, for car drivers of approximately 40 years old in 1985 extending diagonally all the way to car drivers of approximately 70 to 80 years old in 2020. This wave can be attributed to the post World War-II baby-boom of around 1946.
- In the nineties a diagonal downward trough or valley starts to appear on the right of the Figure, in the development of mobility of young car drivers that was almost continuously increasing until that year for car drivers older than about 25 years old. This coincides with the effect of the introduction of the birth-control pill for women around the end of the sixties, as this reduced the number of children born around that time, which thus probably emerges in the mobility figures of car drivers some 20 to 30 years later.

It is interesting to note that these waves and troughs or valleys in the development of mobility run approximately parallel with the progressing age of the car drivers themselves. This cannot be inferred from Figure 4.4, which suggests that there are no serious cohort effects in the developments of the individual mobility of car drivers over time.

Summarising, we have seen that annual population figures disaggregated by age year allow us to derive the annual distance travelled for each age *year* from annual mobility data issued in age *classes* (as was done in the Dutch MON survey). The annual population figures of Statistics Netherlands disaggregated by age year are also found to be very helpful in obtaining adequate forecasts for the distance travelled per age year, both in and out of sample (using data up to 2007).

All results discussed in the present section provided the impetus for deciding to use the population forecasts issued by Statistics Netherlands in order to obtain improved forecasts for future developments in mobility per age class and gender. The central idea is that, ultimately, these improved mobility forecasts will help to obtain better forecasts for victim counts disaggregated by age and gender as well, as will be discussed in the next section.

4.3 Forecasting victim counts per age class and gender

The results presented so far in this Chapter indicate that the use of population figures may be an enrichment of mobility data. If it can be known that the per capita mobility is constant for drivers of every age, it would even be possible to use population data as a substitute, or proxy, for distance travelled. Of course, this is not generally the case.

When (small) age classes are considered, important effects due to development of population figures need to be considered. The examples in Figure 4.1 and Figure 4.2 are clear demonstrations of the potential importance of the development of population figures for explaining the development of travel.

It is also clear from the examples in Figure 4.1 and Figure 4.2 that population figures do not perfectly predict distance travelled. This is due to the inaccuracy of the data on distance travelled alone, or due to economic developments that may cause non-linear or sudden changes in the per capita mobility. It may be possible to build a model for individual mobility, or obtain such information from separate research. A combination of such information and population figures may then be used as a substitute for mobility information. In the future, this line of research may produce improvements over the models presented in this section, but this is however not considered in this section.

In this section we assess the importance of the use of population figures for forecasts, in addition to (or as a substitute for) mobility data. It is clear that valuable mobility information, should never be replaced by population data. However, in some cases no reliable mobility information or forecast is available.

To this end, basically, three different models were fitted: a model using mobility data only, a model using mobility and population size data and finally a model using population size data only. The latter two models exploit the finding of the linear development of the mobility per capita in the previous section. The results of these three model types are used to assess the general importance of the use of population figures.

It is found that generally, long term forecasts using population figures only is roughly similar to forecasts using mobility information and population figures, considering the fact that a long term forecast is carried out. In many cases, forecasts using mobility information only, yield substantially different results. It appears not possible to (fully) explain such differences from subtle differences in the models. It seems likely that forecasts based on exponential trend are not capable of describing (future) changes in mobility due to the changes in population, as a consequence of e.g. the babyboom or the birth control pill.

In order to reduce the possibility of the results just being coincidental, the analysis is performed for many different model assumptions (including interventions), analysed for both the number of fatalities and the number of persons killed or seriously injured (KSI) in road accidents, different numbers of age groups and different age groups. The approach described in this section assumes the availability of (a prognosis) of population figures, and assumes no separate prognosis is available for the mobility information of the age groups considered.

4.3.1 Technical details of the model

In all cases log-linear multivariate structural times series models were used (for standard text books discussing these models see Harvey (1989), Durbin and Koopman (2001), and Commandeur and Koopman (2007), for the specific multivariate extension used here, see Bijleveld et al. (2008)), which were fitted with Ox (Doornik, 2001) and SsfPack (Koopman, Shephard, and Doornik (1999)). Such models are also discussed in D7.7. Technical details are provided in Appendix A4.3.

Annual Dutch data on the number of fatalities or on the number of persons killed or seriously injured (KSI) in road accidents for the years 1985-2007 were used, as well as on distance travelled according to the MON survey, disaggregated by age and gender. In addition annual data on the size of the Dutch population disaggregated by age and gender were used,

The following divisions into age classes were considered:

- two age classes (both 0-49 year old road users versus 50+ year old road users, and 0-24 old road users versus 25+ old road users),
- three age classes (road users of 0-19 years old, 20-64 years old, and 65+ years old),
- four age classes (road users of 0-14 years old, 15-24 years old, 25-64 years old, and 65+ years old),
- and five age classes (road users of 0-14 years old, 15-19 years old, 20-29 years old, 30-64 years old, and 65+ years old).

Finally, we had to give some special attention to the sudden drop in the number of fatalities in The Netherlands in 2004 (Stipdonk et al, 2004). Therefore, all these analyses were performed using three forecast structures:

- forecasts until 2040 from the year 2004 onwards,
- forecasts until 2040 from the year 2007 onwards excluding a level shift intervention variable for the risks in 2004, and
- forecasts until 2040 from the year 2007 onwards including a level shift intervention variable for the risks in 2004.

Structural changes in road safety conditions may have occurred near the end of the observed period. This reduces the viability of a validation of the model by simulating prognosis, e.g. because failure need not be attributable to the model.

4.3.2 Highlights of the resulting forecasts

All combinations of age groups and forecast structures obviously resulted in a large amount of results, and it is impossible to present and discuss them all in this paper. We will therefore describe here only some examples of the most important findings of this analysis. More highlights of the results can be found in Appendix A4.4.

First an example of how the development in time of the number of victims may strongly depend in age, gender and the way data are modelled:

- The forecast of the number of male victims between 20 and 64 years of age is given for two models: one for data until 2003, and one for data until 2007. The sharp drop in the number of male casualties, especially in this age group, generates different forecasts.
- By contrast, the forecast for females KSI of 0-19 hardly depends on the choice of the period the forecast is based on.. Clearly the drop in casualties in 2004 does not apply to every subgroup of road users.

The effects of including data on population size in the estimation of disaggregated forecasts until 2040 also gives several important results:

- First of all, the inclusion of population figures in the analysis is especially relevant for the forecasts of males in the 20-64 age groups, and for the males and females in the 65+ age groups, since it is for these subgroups that the largest differences in forecasts between the three models are found. By including the population data, mobility forecasts are corrected for population changes, and therefore can be believed to be more realistic.
- There were significant drops in the fatalities in 2004 for the males of 0-19 years old and 20-64 years old, but not for the other four subgroups. The inclusion of a level shift intervention variable for the risk in 2004 is especially relevant for the male fatalities in the 20-64 years group, since the forecasts including and excluding intervention variables in 2004 are markedly different for this subgroup.
- Finally, only the males of 0-14, 15-24 and 25-64 years old and the females of 15-24 years old display significant reductions in fatalities in 2004.

4.4 Conclusion

This Chapter investigated the relevance and effects of including data on population size in time series analysis of developments in mobility, risk, and accident outcomes disaggregated by age and gender. The main objective of this investigation was to ultimately obtain better forecasts for these subgroups of road users.

The inclusion of population figures in these analyses was motivated by the fact that the development of the population disaggregated by age is found to display a number of distinct waves and troughs over time. Since the development in the age composition of the population affects the development of the mobility of age years and groups of the road user population, the incorporation of such waves and troughs extending into the future appear extremely relevant for forecasting mobility figures disaggregated by age and gender, and are therefore helpful in

obtaining better forecasts for road accident outcomes as well. When no prognosis for mobility is available that accounts for such waves and troughs in the development of travel, population figures may be a reasonable substitute. It should be noted that it is impossible to formally quantify the reliability of prognosis.

It appears that cohort effects can be ignored for main transport, and that mobility is quite well described by population size and a linear trend describing individual distance travelled.

The results found in Section 4.3 show a remarkable similarity among models including population figures. Differences become apparent in the distant future, which spans more than 30 years in the graphs. The validity of the models at that range is however highly questionable. Differences with prognosis from the model without population figures, in particular those found in the short term are most likely due to population effects rather than subtle model differences.

Finally it is noted that modelling started with even further detailed disaggregations where age and gender subgroups are considered for separate collision opponents (car-car, car-bicyclist, et cetera) showing promising results. However, modelling complexity mounts if only due to the substantially increased number of age and gender combinations to be considered. It is therefore needed to first devote some research into which specific cohorts to consider, possibly separately for collision opponent and victim (driver).

4.5 References

Bijleveld, F.D., Commandeur, J.J.F., Gould, P.G., & Koopman, S.J. (2008). Model-based measurement of latent risk in time series with applications. *Journal of the Royal Statistical Society Series A* 171(1), pp. 265-277.

Commandeur, J.J.F. & Koopman, S.J. (2007). *An introduction to state space time series analysis*. Oxford: Oxford University Press.

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

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

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

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

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

Stipdonk, H.L., Aarts, L.T., Schoon, C.C. & Wesemann, P. (2006). De essentie van de daling in het aantal verkeersdoden. Ontwikkelingen in 2004 en 2005, en nieuwe prognoses voor 2010 en 2020. (The essence of the decrease in the number of road traffic fatalities. Developments in 2004 en 2005, and new forecasts for 2010 and 2020). SWOV report R-2006-4. Leidschendam: SWOV.

Van den Bossche, F.A.M. (2006). *Road Safety, Risk and Exposure in Belgium: An econometric approach*. Dissertation. Hasselt University, Diepenbeek, Belgium.

Van den Bossche, F.A.M., Wets, G. & Brijs, T. (2007). Analysis of road risk by age and gender category. Time series approach. *Transportation Research Record. Journal of the Transportation Research Board*, 2019, pp. 7-14.

Chapter 5 - Modelling meteorological effects on road safety

Yvette van Norden (SWOV), ed.

5.1 General introduction

Weather conditions are generally considered as a relevant factor affecting the number of road accidents and associated casualties in a given network. On the other hand, weather affects mobility as well. Moreover, the influence of weather on mobility differs according to the type of network, whether motorways, rural roads or urban roads. According to relevant literature (see for example Hermans et al. (2006)), weather may explain in the order of 5% of the accident/fatality variability. Therefore, models that include only weather effects as explanatory variables cannot be expected to have predictive value. Instead, these models can be expected to provide some insight into the expected effects of weather conditions both in the form of type of effect (positive/negative) and relative magnitude.

In two recent publications, some interesting results were found. Hermans et al. (2006) developed state-space models at a national level for Belgium for the period 1974-1999. It was concluded that the quantity of precipitation (in mm, measured as an average for the whole country) was only associated to an increase in the number of accidents and casualties of moderate severity. Eisenberg (2004) also noticed that under adverse weather conditions drivers possibly adopt lower speed resulting, on average, to less severe accidents.

The influence of temperature and rainfall on the aggregate numbers of injury accidents and victims on a monthly basis were first studied by Scott (1986) by modelling the monthly number of injury accidents in the UK from 1970 to 1978. Since then, the climatic factor was integrated in the systemic approach of aggregate risk modelling, (Lassarre, 1984), (Gaudry, Lassarre, Eds, 2000) and in particular (Fridstrom, 1995) introduced meteorological information including both average and atypicality within the month. Aggregate results on French datasets were produced on a daily basis (Bergel et al, 1995) and on a monthly basis using both average and atypicality within the month (Bergel, Depire, 2004). Studies which involve mobility data as well are (Bergel, 1992), (Jaeger, 1998) and more generally also (Gaudry, Lassarre (Eds), 2000) in which results related to the weather effects on the mileage in different countries are presented. References of estimation of weather effects on road demand and road risk at an aggregate level are given in (Bergel, 2008).

In this Chapter, two studies on the effect of weather conditions on the road safety are presented. In Section 5.2, the impact of temperature and precipitation on the number of injury accidents and fatalities in the Athens area is discussed. Some results on the number of injury accidents in France, the Netherlands and the region of Athens in relation to three weather variables (temperature, rainfall

and frost) are presented in Section 5.3, including a comparison between the three different regions.

5.2 Impact of meteorological variables on the number of injury accidents and fatalities in the Athens region

George Yannis, Constantinos Antoniou, Dimitris Katsochis, Eleonora Papadimitriou (NTUA)

5.2.1 Introduction

This section is aimed at the exploration of the impact of two essential meteorological indicators (temperature and precipitation) on the number of total accidents and fatalities recorded in the wider Athens area for a period of up to 20 years. The suitability of several distributions for modelling road safety data that are temporally correlated is investigated. Based on the findings of previous research (Yannis et al., 2007, Hermans et al., 2006), the negative binomial distribution is chosen. More precisely, the correlation between accidents and fatalities, and meteorological variables is examined. Exposure data (in the form of vehicles recorded in a regional toll station) is also included in the analysis.

The analysis involves different methodological approaches, including Generalized Linear Models (GLM) and Dynamic Generalized Linear Models (DGLM) or state-space models. In the former case, two models have been developed, one for the total number of accidents and one for the number of fatalities, while in the latter only one model for accidents. Different approaches provide insight into the modelled processes and demonstrate their use.

In Subsection 5.2.3, models including only weather effects as explanatory variables are used to provide some insight into the expected effects of weather conditions on the number of accidents and the number of fatalities. Further models including exposure data, which indeed have further explanatory and especially predictive capacity, are presented in Subsection 5.2.4.

5.2.2 Data used

Average temperature and total precipitation are available at daily level for the wider Athens area since 1985 from the National Observatory of Athens (NOA) and the National Statistical Service of Greece (NSSG). The dataset that is used in this analysis covers the period until the end of year 2005. Data have undergone some processing from the Department of Transportation Planning and Engineering (School of Civil Engineering) of the National Technical University of Athens (NTUA).

Furthermore, in an attempt to also consider exposure data, monthly traffic data have been used for the period from January 1997 to December 2005 (108 months). The values that have been used refer to the Shimatari toll station to the north of Athens, and are assumed as a reasonable proxy to the entire traffic in the Athens area. Information on the traffic composition (passenger cars etc) is also available and is exploited in this analysis.

Figure 5.1 provides an overview of some key variables of the original daily dataset (1985-2005). Periodicity over the years (at monthly level) is evident, especially with respect to temperature and –to a lesser degree- precipitation that are typically reproduced each year.

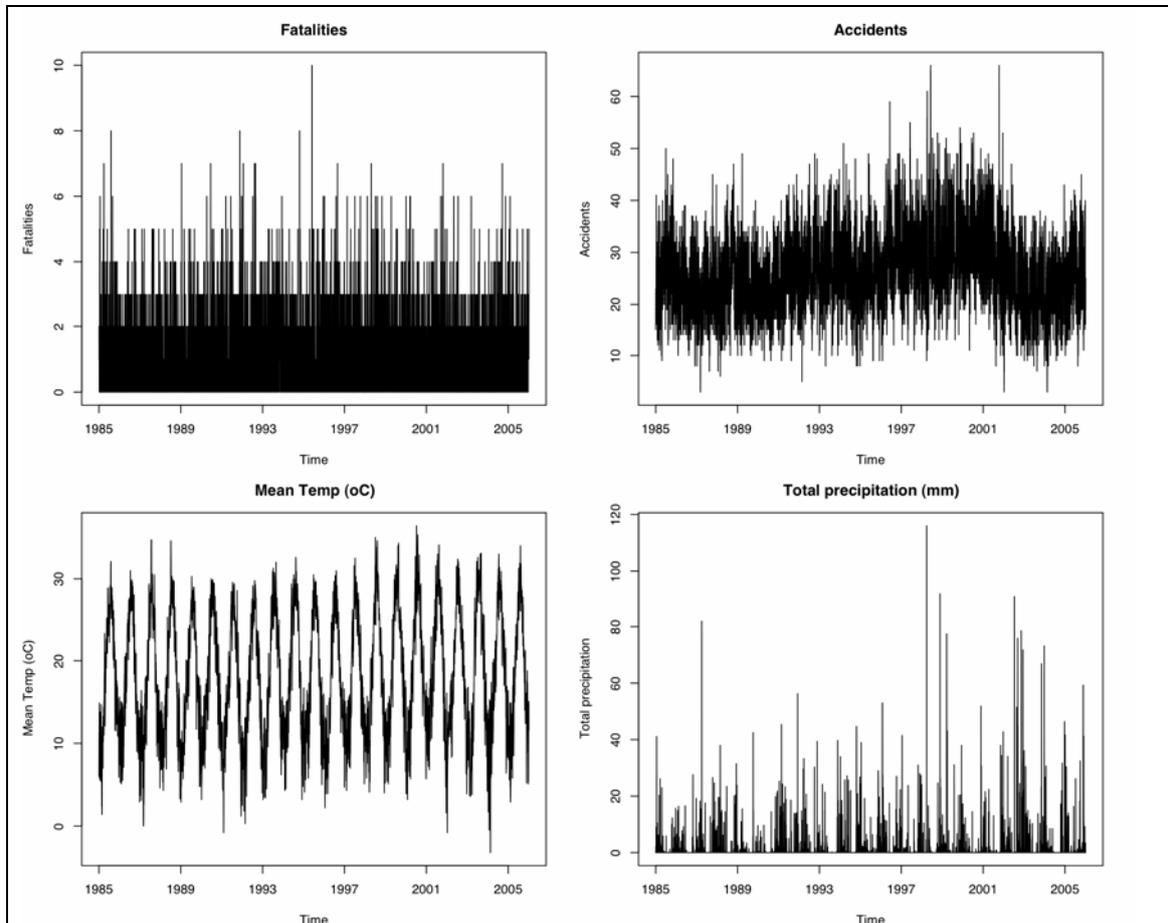


Figure 5.1: Descriptive graphs of the daily dataset used for the analysis

Exposure data was only available at a monthly series and only after 1997. In order to be able to meaningfully incorporate it into the analysis, the original data set (shown also in Figure 5.1) was (a) truncated to only include data after 1997 and (b) was aggregated to monthly level. The data set was also further analysed, so that additional variables that are more appropriate for this analysis were constructed. The main manipulation involved clustering temperature and precipitation values into groups. There is evidence in the literature that meteorological variables may be more useful if they are used as categorical values (as opposed to numerical variables), allowing to isolate e.g. extreme temperatures or days with high precipitation. For example, Laaidi and Laaidi, 2002, found that in France heat waves have strongly positive effects on injury crashes. Possible reasons are that drivers prefer to shift their planned trips to the late evening or late morning. Furthermore, heat waves disturb the sleeping pattern of persons, which could cause a higher degree of tiredness amongst drivers. Figure 5.2 presents the distribution of the original daily weather data.

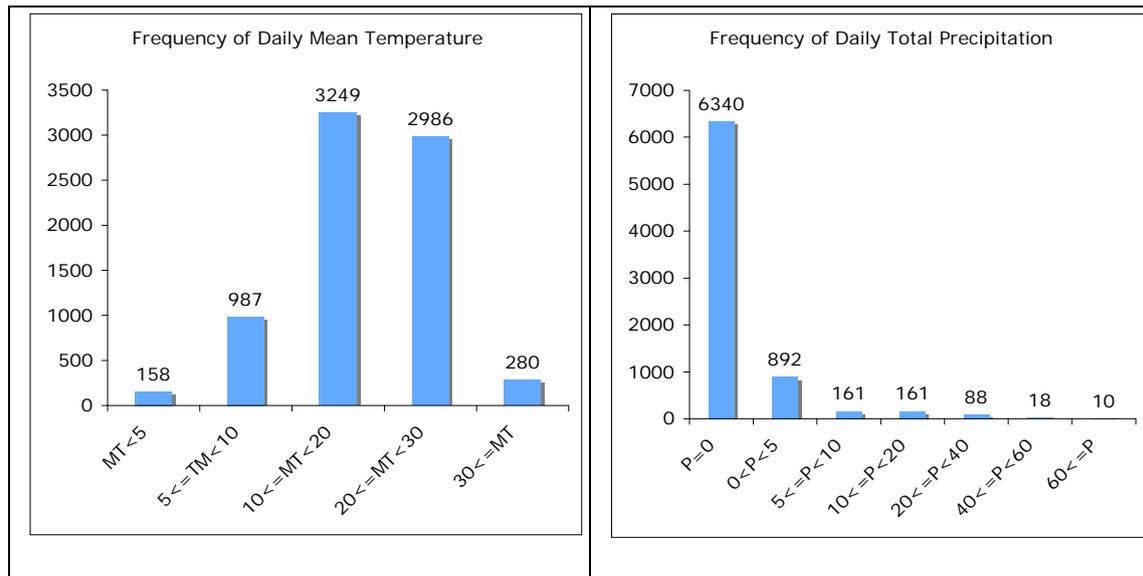


Figure 5.2: Binning of daily weather data

Figure 5.3 presents some key variables of the aggregated monthly data. For the aggregation of the precipitation, a total precipitation for the month was computed. For the aggregation of the temperature, several variables were computed: month average mean temperature, minimum mean temperature for the month, and maximum mean temperature for the month (where mean refers to daily mean). These values were grouped according to the thresholds shown in Figure 5.2.

Several observations can be made. Unusually high precipitation was mainly recorded during winter 2002-03, following two relatively dry winter periods. Data reveal significant improvement (reduction) of fatalities and total accidents. It is stressed, though, that this reduction practically appears as a step in the transition from 2001 to 2002. On the other hand, a steady increase of passenger car traffic is recorded at annual level.

The proportion of passenger cars of the total number of vehicles travelling through the aforementioned toll station appears to increase over time. In particular, cars amounted up to approximately 75% of all vehicles for the first three years (1997-99). This proportion increased to 80% in 2000, 85% for the period 2001-2004 and about 90% in the second half of year 2005.

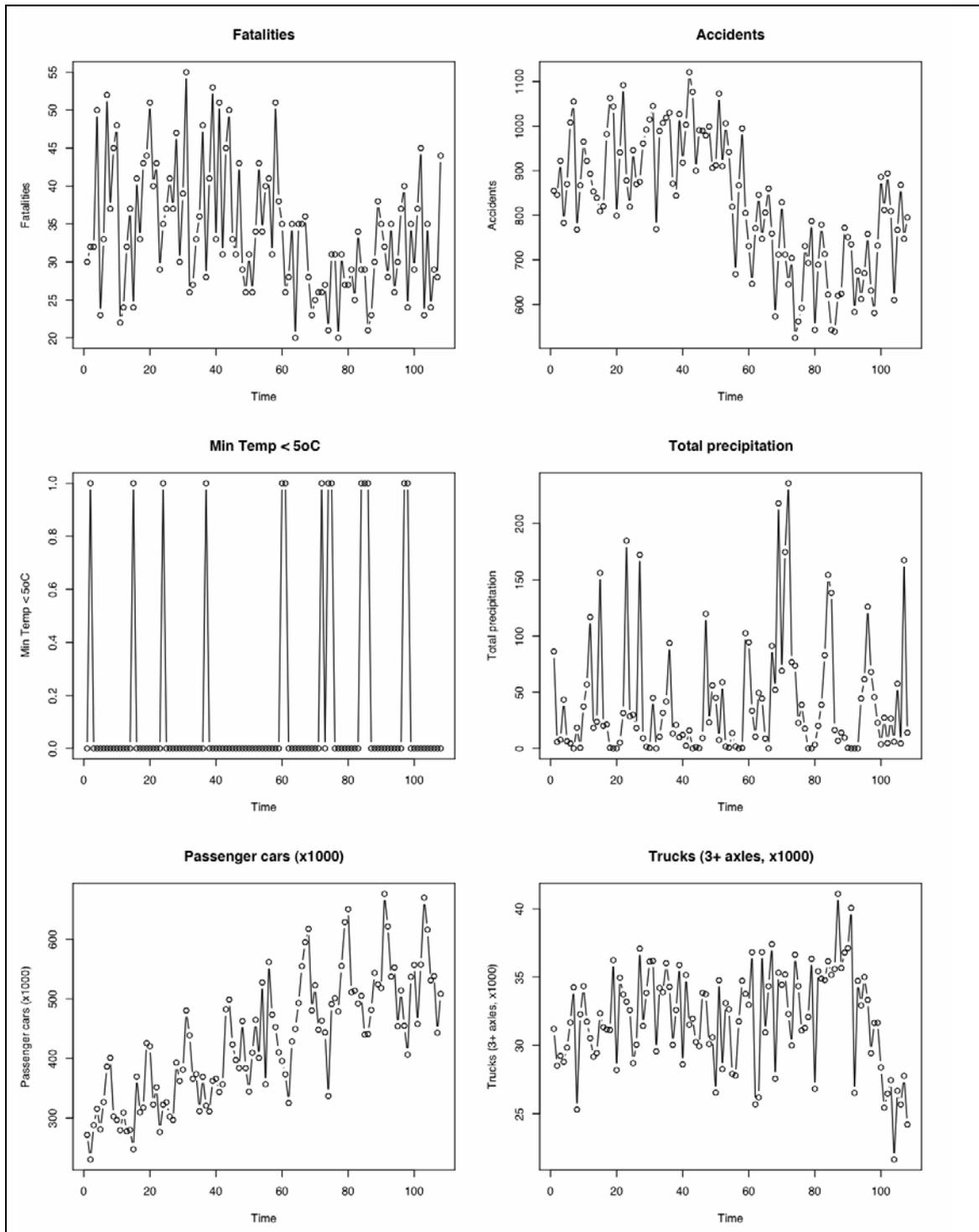


Figure 5.3: Descriptive graphs of the monthly dataset used for the analysis

5.2.3 Daily analysis without exposure data

Working with different data sources inevitably leads to mismatches in the granularity and scope of the data. In this research, meteorological and accident data are available at a much higher resolution (i.e. daily) and span a much larger time period (1985 to 2005). On the other hand, exposure data are only available for the last 9 years and only at a monthly level. In order to both explore the detailed meteorological data and obtain some insight into the effects

of traffic, two different analyses have been made: one with the meteorological data alone (using daily time series data between 1985 and 2005) and one with the aggregated time series (used monthly data from 1997 to 2005). This analysis has been repeated for both accidents and fatalities. The distributional assumption that is made is that of negative binomial, which has been shown to outperform other distributions in similar models.

Table A5.1 (see Appendix A5.1 for all numerical information in tabular form) presents the estimation results for accidents using the daily data set. The main findings related to weather effects are that the increase of the temperature leads to an increase in the number of accidents, which is similar to the findings in the literature, e.g. Hermans et al. (2006). On the other hand, an increase of precipitation leads to a decrease in the number of accidents, possibly due to lower overall traffic and extra care exercised by the drivers. Both trends are consistent, since the coefficients increase in absolute terms while moving to clusters with higher value ranges. Furthermore, a lower number of accidents are observed in August, presumably reflecting the lower traffic flow during that month (note that this data covers the Athens region, which has very low traffic during August) due to summer holiday leaves.

As mentioned in Subsection 5.2.2, the two meteorological variables are grouped to yield clusters of temperature ranges. This approach has several advantages, including the ability to model non-linear patterns. In each case, the first group serves as the base of comparison (i.e. the number of days with mean temperature below 5 °C and with total precipitation less than 5 mm respectively).

Table A5.2 presents the respective model estimation results for the fatalities. A similar effect is observed for the temperature, higher values of which are associated with higher fatality numbers. In terms of precipitation, it was found that very high precipitation values were associated with a decrease in the number of fatalities. As stated in the general introduction, meteorological variables alone do not possess sufficient explanatory power to completely describe the examined phenomenon, so trying to make predictions on such models is not practical.

5.2.4 Results

The analysis of the previous subsection is based on the theory of generalized linear models. In this section, the approach of dynamic generalized linear models (DGGLM), which allows for –among other things- explicitly modelling serial correlation, is used instead. While state-space models (SS) and the traditional Kalman Filter techniques have normality requirements, this family of models allows distributions of the measurement equation that fall within the exponential family (Dethlefsen & Lundbye-Christensen, 2006).

The DGGLM models are a certain form of SS models and were run at this context using the Poisson distribution and log link for the dependent variable (using the `sspir` package in R) and as a GLM (again Poisson with a log link function). This is considered a standard approach for the analysis of road safety data

according to the literature (e.g Nicholson and Wong, 1993, Yannis et al., 2007), due to the suitability of the Poisson distribution to describe counts data.

The following explanatory variables have been used:

- MinMeanTempLT5: binary (0/1) variable taking the value 1 (or TRUE) if the minimum mean temperature of at least one day in the month was less than 5 degrees C.
- SumTotalPrecipitation: sum of total precipitation (in mm) in month
- HeavyTrucks: number of heavy trucks (three axle or more)
- Motor2Wheelers: mopeds & motorcycles

Table A5.3 provides model estimation results for the monthly data set, including exposure data. It can be seen that very low temperature values (minimum daily temperature for the month less than 5 °C) result in a reduction of accidents, possibly due to the lower traffic. Higher total precipitation also results in lower accident rates. Furthermore, higher passenger car traffic results in a lower accident number (presumably due to the lower speeds that are allowed due to higher congestion), while a higher number of heavy trucks leads to an increase in accidents –in fact, the latter goes with a 10-times larger coefficient.

In the SS model, the intercept is allowed to vary, and there is an unstructured seasonal pattern (as described in the relevant paper by C. Dethlefsen & S. Lundbye-Christensen). Therefore, the differences between the two models are mainly focused at the time-varying intercept in the state-space model and the unstructured seasonal pattern component.

Figure 5.4 shows the observed, estimated and predicted data for the monthly accident data set. Note that the data used for the estimation did not include the period used for prediction (2005).

Already from a visual inspection of the models' results, it is clear that the SS performs better than the GLM. However, in order to also provide a more rigorous check on the validity of this argument, the RMSPE goodness-of-fit metric has been computed to show each model's estimating/predictive ability (see Table 5.1).

The results of the DGLM for the accidents model may be found in Table A5.3 of Appendix A5.1. It is interesting to note the different signs of the variables for the presence of heavy vehicles or powered-two-wheelers (PTW), which is reasonably explained by the different size and weight of those two vehicle classes. It is also worth noting that the coefficient for trucks is five times larger than the one for PTW.

	Estimation (1997-2004)	Prediction (2005)
SS	0.0386	0.0461
GLM	0.0727	0.0984

Table 5.1: Estimation and prediction performance of GLM and SS models

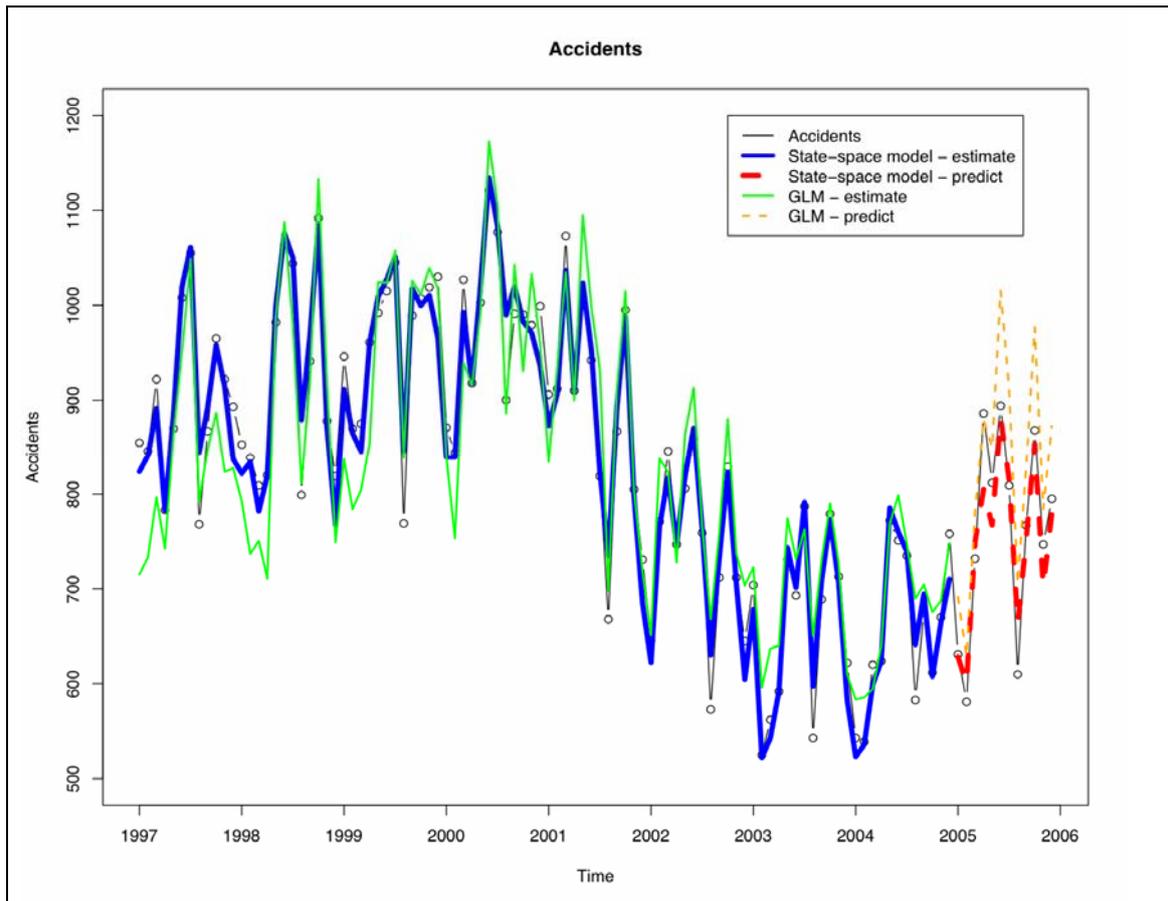


Figure 5.4: State-space and GLM model estimation results for accidents

5.2.5 Conclusion

The results in this section may be considered encouraging, as there are consistent trends regarding the impact of temperature and precipitation on the examined dependent variables.

These rather simple models demonstrate a reasonable differentiation across months within a year, with June yielding more accidents than each month of the autumn period, probably because more vehicle-km are driven on most road networks during early summer. Intuitive expectations seem to be justified to some extent from the use of models that also utilise exposure data. In particular, it appears that low temperature during wintertime, mostly, corresponds to some reduction of recorded accidents. The same is the case as total precipitation in a month increases, probably due to reduced mobility under rainy weather, but this effect is much less pronounced.

5.3 Modelling the weather effects on the numbers of injury accidents at an aggregate level in different regions

**Ruth Bergel, Mohammed Debarh (INRETS),
With the collaboration of Frits Bijleveld, Niels Bos (SWOV) and Eleonora Papadimitriou (NTUA)**

5.3.1 Introduction

In France, for the period 1990-2000, 14% of the injury accidents were registered by rainy weather, and 1% only, at the most, in case of fog, in case of glaze, or in case of snow/hail (ONISR, 2006). Among these factors, fog causes accidents with the highest severity level (accidents due to fog are twice as severe than the average), specially on motorways, due to the lack of visibility, Rain, on the contrary, causes accidents which are a little less severe than the average, but which are numerous. The stake, measured in number of injury accidents, has led to consider rain as a major meteorological factor explanatory of the road risk (Brodsky, Hakkert, 1988).

The aim of this section is to take into account three weather variables, which influence both mobility and road safety, in the analysis of the development of the numbers of injury accidents, aggregated at a national level and according to the network type, on a monthly basis for a long observed period. Aggregate datasets of numbers of injury accidents were gathered for France, for the Netherlands and for the region of Athens. The same variables were used in each country/region, which measure: rainfall, temperature and the occurrence of frost. Their global effects on the numbers of injury accidents were estimated via exogenous parameters, and compared. Several variables were used as measures of exposure to risk at an aggregate level, with the objective of quantifying both the direct weather effects on the risk indicator, and their indirect effects through mobility variations. We shall focus on the significance of the weather effects, in magnitude and in sign, on a monthly basis, and we shall provide an interpretation of the estimated parameters highlighting these weather effects and give some orders of the size of their magnitude.

5.3.2 Methodology and data

State space methods are used in this section, and more precisely the structural time series analysis with explanatory variables (Harvey, 1989), (Commandeur and Koopman, 2007). The methodology (econometric specification and related packages) and data used are described in Appendix A5.2. The data used where in all cases monthly data, in several cases averaged or aggregated from daily data. For each of the three considered regions long time period data are available: 1975-2000 for France, 1987-2005 for the Netherlands, and 1987-2005 (or only 1997-2005 in case exposure data were included in the analysis) for the region of Athens. The risk variable was in all cases the number of injury accidents, aggregated at a national level, or disaggregated according to the type of network. The meteorological factors retained were measured with three weather variables, defined at a daily level:

- the maximum temperature of the day (1/10 °C)
- the occurrence of frost (no or yes)
- the rainfall height (mm)

For France and for the Netherlands, daily climate variables were first calculated by averaging daily variables measured at meteorological observation points spread over the whole territory. For the region of Athens, the daily variables were measured at the observation point of Hellenika, in the south of Athens. These variables were then aggregated or averaged in order to construct the monthly variables.

Figures A5.1-A5.3 in Appendix A5.2 show the development of the monthly weather variables used for France, for the period 1975-2000. Clearly, differences between the seasonal patterns (i.e. the periodic movement within the year) of different years appear, in particular for the temperature and the occurrence of frost. For instance, during the cold winters of the years 1985 to 1987, the maximum temperature of the day (averaged over a hundred meteorological stations and over the month) was only 2-3 °C for the months of January/February, whereas the seasonal average is 7-8 °C for these months. The number of days of frost (the occurrence of frost, averaged over the hundred stations and over the month) reached 20-23 at the same period, whereas the seasonal average is 12-13 for these months. At last, it can be seen that the amount of precipitation, also subject to more variability than the other variables, was higher than usual during the year 1991.

These differences to an average seasonal pattern - in other words these differences to the seasonal average - constitute indeed the information of interest, as their influence on the number of injury accidents cannot be captured in the seasonal component in a time series analysis of the observed series.

Proxies of unavailable variables were used for measuring exposure to risk at an aggregate level: oil sales for France, survey data for the Netherlands, and vehicle counts at the Shimatari toll station at the north exit for Athens. On the French motorways and main roads on the contrary, the number of vehicle-kilometres registered, measured with accuracy, was used.

5.3.3 Results

Model Fit and Diagnostics

The explanatory variables were all retained in the model, whether significant or not, in order to compare their significance within the same model, or among different similar models. For each risk indicator (corrected for the explanatory variables effects), the level, slope and seasonal component were provided for the whole period, and the residuals. The variances of these four components were estimated, and tests of hypothesis of the nullity of the level, slope and seasonal components variances were performed. The usual tests of validity of the residuals were performed, and were satisfactory.

The numerical results related to the structure (error variances) and the regression exogenous coefficients are given for each model, in tables A5.4 to A5.13 in Appendix A5.2.

Global weather effects

The following comments are limited to the significant parameters (at a 5% significance level).

At a national level, the global weather effects on the number of injury accidents are coherent for France and for the Netherlands.

- Rainfall is positively correlated to the number of injury accidents (100 mm of additional rainfall during a month results in an increase of 0.2%-0.3% in the number of injury accidents that month)
- Temperature is positively correlated to the number of injury accidents (1° C of additional average temperature during a month results in an increase of 1%-2% in the number of injury accidents that month). Note that for example, a month in which the temperature is a bit higher on every day of the month, and a month with only a very few days with a much higher temperature may both have in the same increase for the average temperature during the month. But the influence on the number of injury accidents will probably be different. So, this result also depends on the period over which the temperature is averaged, in this case average temperature per month.
- The occurrence of frost is negatively correlated to the number of injury accidents (1 additional day of frost during a month results in a decrease of 0.3%-0.6% in the number of injury accidents that month).

At a specific network's level, these effects happen to differ between the three observed regions, in magnitude and even in sign:

- Rainfall is positively correlated to the number of injury accidents on motorways in France and the Netherlands (100 mm of additional rainfall during a month results in an increase of about 0,5% to the number of injury accidents on main roads and on motorways in France and in an increase of 0,8% in the number of injury accidents on motorways in the Netherlands). The only exception is that rainfall is negatively correlated to the number of injury accidents in the region of Athens, in major part constituted of urban roads (100 mm additional rainfall during month results in a decrease of about 0.5 % of the number of injury accidents),
- Temperature is positively correlated to the number of injury accidents, but with different magnitudes on main roads and motorways in France (1°C of additional average temperature during a month results in an increase of 0,4% and about 2% of the number of injury accidents on main roads and motorways respectively), on rural roads and motorways in the Netherlands (1°C of additional average temperature during a month results in an increase of about 1% and 2%-3% of the number of injury accidents, on rural roads and motorways in the Netherlands respectively). In the region of Athens, on the contrary, the impact of temperature is only significant in the winter (1°C of additional average temperature during a month results in an increase of 2% in the number of injury accidents).
- At last, the occurrence of frost, does not appear to have a significant effect on the number of injury accidents on neither of the two considered networks, whether for France or for the Netherlands.

Direct and indirect weather effects

These global weather effects on the level of the number of injury accidents were registered under variations of mobility. It would be interesting to try to decompose these effects into two components:

- the *direct* weather effects on the number of injury accidents,
- the *indirect* effect, where mobility changes are induced by the weather

To decompose weather effects into direct and indirect components exposure measures are needed. As proxies for the exposure *at the aggregate level*, the following variables were used:

- oil sales for France,
- the number of driver-kilometres provided by survey for the Netherlands
- the number of vehicle registered at the toll station at one exit of Athens),

However, either the variable could not be considered as representative of the true mobility (the elasticity value of the risk indicator with respect to mobility being too small) or the variable was not significant in the model of the risk indicator (as it is the case for the region of Athens). Therefore, it was not possible to decompose the weather effects into direct and indirect components at the aggregate level.

In the case of the two French networks (main roads and motorways), the distance travelled is measured accurately and can be considered as a representative measure of the exposure to risk. So this variable can be used for estimating the direct weather effects on the road risk - the distance travelled being unchanged.

Due to the strong correlation between the temperature and the distance travelled, it is not possible to interpret the coefficients related to the average temperature as temperature effects. This is not the case for the two other weather variables. The most significant result is related to rain: rainfall has a positive direct effect and indirect negative effect, on main roads and on motorways, on the number of injury accidents (100 mm of additional rainfall in the month respectively results in an increase of 0,6 to 0,8% of the number of injury accidents – the distance travelled being unchanged -, and in a decrease of 0,1 to 0,2% - resulting from the decrease of the distance travelled).

5.3.4 Conclusion

The results obtained in this section using structural models under state space form are coherent with the results obtained with other sorts of time series methods. Apart from state space methods, autoregressive methods are classically used for performing time series analyses in road safety research (Dupont, Martensen (Eds), 2007).

The models used in this section are still to be completed by including the variance of the meteorological conditions within the month, exploited on a monthly basis. In order to give in more detail a size of their magnitude, we can refer to the analysis of the meteorological effect on the number of injury accidents provided for France, for the years 1999 and 2000, on a monthly basis, including both types - average and atypicality in the month - of meteorological

information. Although the weather effect, aggregated over the year, is only 1%-2% of the absolute values of the total number of injury accidents, it raises to 10% for specific months (Bergel, 2008).

Similar approaches are to be extended to urban networks, on the one hand, and to victim's indicators (fatalities, hospitalized and light injured) on the other hand. The effect of very low temperatures and of frost in particular, should appear to be more significant on accident's gravity indicators.

5.4 General discussion

As expected, models solely built around meteorological variables only demonstrate limited potential in interpreting trends and may only be used as indicative descriptive tools.

The differences in the models texture and level of detail in the different analyses presented in this Chapter do not necessarily mean that some models are by definition better than other ones. This is mostly related to issues such as data availability and overall context. Different road network users may call for varying data treatment as well. One may say that, in territories where detailed exposure data are not available - and are not expected to be to soon -, there is a necessity to develop more parsimonious models, placing focus on precision in variables selection and modelling technique approach. In such cases there is often some "omitted variables" problem, simply because it is not possible to utilise what is reasonably considered an ideal set of predictors. Still, one may want to obtain some idea of the impact of the remaining parameters to be used. So, if one clearly sets a research objective, it is most probable that even simple tools will be sufficient to reach reliable and applicable results.

The challenge on this research topic is first an improved monitoring of the road safety level. It is crucial to impute the road safety level to the appropriate risk factor. Besides, the magnitude of the weather effects on the road risk level can be reduced, by implementing alert systems in case of risky weather conditions (Aron et al, 2007).

The perspectives of this research area are twofold. The first one is the constitution of national meteorological databases that could systematically be used for monitoring road safety at a national level in Europe. The second one is the constitution of an harmonized measure of exposure to risk at a national level in Europe (Yannis et al, SafetyNet D2.1, 2005). Indeed, it is also necessary to estimate the weather effects on the road risk level - under constant mobility conditions. In the absence of a monthly measure of the number of vehicle-kilometres driven at a national level, provided by means of surveys or counts, an estimation of the monthly oil consumption has been proposed.

5.5 References

Aron, M, Bergel-Hayat, R, Saint Pierre, Guillaume, Violette, E. (2007). Added Risk by Rainy Weather on the Roads of Normandie-Centre Region In France, *11th World Conference in Transport Research (WCTR)*, June 2007, Berkeley, CA.

Bergel R. (1992). *Utilisation de modèles ARIMA avec variables externes dans le secteur des transports*. Proceedings of the 6th World Conference on Transport Research. Lyon.

Bergel R. (2008). *Time series analysis with explanatory variables: applications to transport demand and to road risk*. Doctorat Thesis, University Paris-Est (in French).

Bergel R., Debarh, M. (2008). The main packages available for performing state space modelling. Arcueil.

Bergel R., Depire A. (2004). *Climate, road traffic and road risk – an aggregate approach*. Proceedings of the WCTR'04, Istanbul.

Brodsky H., Hakkert S. (1988). Risk of road accident in rainy weather. *Accident Analysis and Prevention*, Vol.20, N°3, pp 161-176.

Commandeur J.J.F., Koopman S.J. (2007). *An introduction to state space time series analysis*, Oxford University Press.

Dethlefsen, G., and S. Lundbye-Christensen (2006). Formulating State Space Models in R with Focus on Longitudinal Regression Models. *Journal of Statistical Software*, Vol.16, N° 1.

Dupont E., Martensen H.,(Eds) (2007) *Multilevel modelling and time series analysis in traffic safety research - Methodology and Manual*. Deliverables D7.4 and D7.5 of the EU FP6 project SafetyNet, <http://www.erso.eu>

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

Fridstrom L. et al. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis & Prevention*, Vol. 27 n°1, pp.1-20.

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

Hermans, E., Wets, G., Van Den Bossche, F. (2006). Frequency and Severity of Belgian Road Traffic Accidents Studied by State-Space Methods. *Journal of Transportation and Statistics*, Vol.9 n°1, pp.63-76.

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

Jaeger L. (1998). *L'évaluation du risque dans le système des transports routiers par le développement du modèle TAG*. Thèse de doctorat ès sciences économiques, Université Louis Pasteur, Strasbourg.

Laaidi, K. and M. Laaidi (2002). *Météorologie et Sécurité Routière. Via Secura*, Vol. 56, pp. 22-24.

Nicholson, A., and Y-D. Wong (1993). Are accidents Poisson distributed? A statistical test. *Accident Analysis & Prevention*, Vol.25 n°1, pp. 91-97.

ONISR (2005). *The major themes in road safety - rain (results1995-2004)* (in French),
<http://www.securiteroutiere.equipement.gouv.fr/IMG/Synthese/Pluie.pdf>

ONISR (2006). *Road safety in France, Statistics of the year 2005*.

Scott P. P. (1986). Modelling Time-Series of British Road Accident Data. *Accident Analysis & Prevention*, Vol. 18 n°2, pp.109-117.

Yannis G et al. (2005). *State of the Art Report on Risk and Exposure Data*. SafetyNet Deliverable 2.1, <http://www.erso.eu>

Yannis, G, Antoniou, C, Papadimitriou, N. (2007). Road Casualties and Enforcement: Distributional Assumptions Of Serially Correlated Count Data. *Traffic Injury Prevention*, Vol. 8:3, pp. 300-308.

Part III

Appendices

Appendix A2.1

This Appendix discusses the econometric specification used for the structural time series modelling described in Chapter 2, in the general non-seasonal case. By adding stochastic seasonal components to this multivariate local linear trend framework for annual data it can easily be extended to the modelling of quarterly or monthly data.

Define

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

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

Also define

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

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

and

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

with

$$B = \begin{bmatrix} \sigma_{\xi^{(1)}}^2 & \text{cov}(\xi^{(1)}, \xi^{(2)}) & \text{cov}(\xi^{(1)}, \xi^{(3)}) \\ \text{cov}(\xi^{(1)}, \xi^{(2)}) & \sigma_{\xi^{(2)}}^2 & \text{cov}(\xi^{(2)}, \xi^{(3)}) \\ \text{cov}(\xi^{(1)}, \xi^{(3)}) & \text{cov}(\xi^{(2)}, \xi^{(3)}) & \sigma_{\xi^{(3)}}^2 \end{bmatrix}$$

and

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

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

The three-level model discussed in Chapter 2 is a special case of the general state space methods for the analysis of time series (Harvey, 1989; Durbin and Koopman, 2001). In matrix algebra, all state methods can generally be written as

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, H_t) \quad (\text{A2.1})$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim NID(0, Q_t) \quad (\text{A2.2})$$

for $t = 1, \dots, n$, where (A2.1) is called the observation or measurement equation (and ε_t is an irregular component consisting of observation errors or disturbances), and (A2.2) is called the state equation (and η_t contains the state disturbances). The present model for evaluating different levels of road safety is a multivariate local linear trend model.

With the above definitions it can be verified that we obtain the following model. First writing out (A2.1) in scalar notation yields the following three observation equations:

$$\begin{aligned} y_t^{(1)} &= \mu_t^{(1)} + \varepsilon_t^{(1)} \\ y_t^{(2)} &= \mu_t^{(1)} + \mu_t^{(2)} + \varepsilon_t^{(2)} \\ y_t^{(3)} &= \mu_t^{(1)} + \mu_t^{(2)} + \mu_t^{(3)} + \varepsilon_t^{(3)} \end{aligned}, \quad (\text{A2.3})$$

while working out (A2.2) in scalar notation results in the following six state equations:

$$\begin{aligned}
 \mu_{t+1}^{(1)} &= \mu_t^{(1)} + v_t^{(1)} + \xi_t^{(1)} \\
 \mu_{t+1}^{(2)} &= \mu_t^{(2)} + v_t^{(2)} + \xi_t^{(2)} \\
 \mu_{t+1}^{(3)} &= \mu_t^{(3)} + v_t^{(3)} + \xi_t^{(3)} \\
 v_{t+1}^{(1)} &= v_t^{(1)} + \zeta_t^{(1)} \\
 v_{t+1}^{(2)} &= v_t^{(2)} + \zeta_t^{(2)} \\
 v_{t+1}^{(3)} &= v_t^{(3)} + \zeta_t^{(3)}
 \end{aligned} \tag{A2.4}$$

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

Appendix A2.2

The estimated variance matrices of the observation disturbances, and of the state disturbances, estimated for the datasets of the whole of France and the whole of the Netherlands, with Ox/Ssfpack, are given here.

	oil sales	injury accidents	fatalities
oil sales	9.0573E-5	15.5929E-5	13.5539E-5
injury accidents	15.5929E-5	59.6208E-5	44.5838E-5
fatalities	13.5539E-5	44.5838E-5	91.1089E-5

Table A2.1: Estimated variance matrix of the observation disturbances for France

	exposure	accident risk	fatality risk
<i>level components</i>			
exposure	3.68091E-5	-1.03146E-5	3.61062E-5
accident risk	-1.03146E-5	31.8722E-5	14.1376E-5
fatality risk	3.61062E-5	14.1376E-5	10.8083E-5
<i>slope components</i>			
exposure	7.01823E-7	-1.12504E-6	1.0619E-6
accident risk	-1.1250E-6	1.80348E-6	-1.70236E-6
fatality risk	1.06196E-6	-1.70236E-6	1.060691E-6
<i>seasonal components</i>			
exposure	3.98743E-7	-2.2089E-7	1.48649E-6
accident risk	-2.2089E-7	1.22366E-7	-8.23466E-7
fatality risk	1.48649E-6	-8.23466E-7	5.54155E-6

Table A2.2: Estimated variance matrices of the state disturbances for France

	motor vehicle kms	injury accidents	fatalities
motor vehicle kms	24.1432E-5	18.4007E-5	54.0755E-5
injury accidents	18.4007E-5	161.536E-5	203.46E-5
fatalities	54.0755E-5	203.46E-5	821.25E-5

Table A2.3: Estimated variance matrix of the observation disturbances for The Netherlands

	exposure	accident risk	fatality risk
<i>level components</i>			
exposure	2.67605E-5	4.02208E-5	-8.8639E-5
accident risk	4.02208E-5	0.28.5414E-5	-0.23.0567E-5
fatality risk	-8.8639E-5	-0.23.0567E-5	0.000335721
<i>slope components</i>			
exposure	9.41423E-8	-2.70138E-7	1.46211E-7
accident risk	-2.70138E-7	7.75149E-7	-4.19546E-7
fatality risk	1.46211E-7	-4.19546E-7	2.27078E-7
<i>seasonal components</i>			
exposure	6.46107E-6	-2.26173E-5	-8.50118E-6
accident risk	-2.26173E-5	7.91729E-5	2.97588E-5
fatality risk	-8.50118E-6	2.97588E-5	1.11855E-5

Table A2.4: Estimated variance matrices of the state disturbances for The Netherlands

Appendix A3.1

This section presents all Tables that correspond to the Figures obtained for each examined country. The Tables include information on the breakpoints estimates and parameter coefficients.

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.BEmot	324.0	11.51	(1 st break point)	
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	8.26	1.02	8.13	-
BEmot	0.03	0.004	7.18	# left slope
U1.BEmot	-0.06	0.005	-11.59	# diff. in slopes

Null deviance: 284.0 on 45 degrees of freedom
 Residual deviance: 60.2 on 42 degrees of freedom
 AIC: 152.91

Table A3.1 Model output for Belgium

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.GRmot	335.30	11.49	(1 st break point)	
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	7.50	0.34	22.2	-
GRmot	0.04	0.002	20.1	# left slope
U1.GRmot	-0.07	0.005	-15.1	# diff. in slopes

Null deviance: 925.0 on 45 degrees of freedom
 Residual deviance: 61.2 on 42 degrees of freedom
 AIC: 153.65

Table A3.2 Model output for Greece

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.NLmot	219.3	3.8	(1 st break point)	
psi2.NLmot	359.3	5.7	(2 nd break point)	
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	12.09	0.75	16.05	-
NLmot	0.06	0.005	11.46	# first slope
U1.NLmot	-0.17	0.007	-24.47	# slope 2 – slope 1
U2.NLmot	0.08	0.006	14.29	# slope 3 – slope 2

Null deviance: 1745.1 on 45 degrees of freedom
 Residual deviance: 24.9 on 40 degrees of freedom
 AIC: 116.33

Table A3.3 Model output for the Netherlands

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.ESmot	61.05	10.60		(1 st break point)
psi2.ESmot	231.20	20.43		(2 nd break point)
psi3.ESmot	262.20	13.29		(3 rd break point)
psi4.ESmot	310.70	80.27		(4 th break point)
psi5.ESmot	375.90	16.22		(5 th break point)
psi6.ESmot	452.70	17.72		(6 th break point)
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	4.90	1.39	3.53	-
ESmot	0.17	0.04	4.19	# first slope
U1.ESmot	-0.15	0.04	-3.63	# slope 2 – slope 1
U2.ESmot	-0.12	0.11	-1.09	# slope 3 – slope 2
U3.ESmot	0.17	0.13	1.33	# slope 4 – slope 3
U4.ESmot	0.04	0.09	0.39	# slope 5 – slope 4
U5.ESmot	-0.24	0.08	-3.08	# slope 6 – slope 5
U6.ESmot	0.11	0.06	1.90	# slope 7 – slope 6

Null deviance: 649.8 on 45 degrees of freedom
 Residual deviance: 82.2 on 32 degrees of freedom
 AIC: 187.26
Table A3.4 Model output for Spain: Complex model

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.ESmot	357.70	22.34		(1 st break point)
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	10.65	0.76	13.96	-
ESmot	0.03	0.004	7.63	# first slope
U1.ESmot	-0.07	0.01	-7.36	# slope 2 – slope 1

Null deviance: 649.8 on 45 degrees of freedom
 Residual deviance: 223.0 on 42 degrees of freedom
 AIC: 213.15
Table A3.5 Model output for Spain: Simplified model

Estimated Break-Point(s):				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>		
psi1.CZmot	149.3	7.5		(1 st break point)
psi2.CZmot	274.6	5.0		(2 nd break point)
psi3.CZmot	319.5	7.5		(3 rd break point)
Meaningful coefficients of the linear terms:				
<i>Notation</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Slope</i>
(Intercept)	4.67	2.39	1.96	-
CZmot	0.07	0.02	3.49	# first slope
U1.CZmot	-0.12	0.02	-5.68	# slope2 – slope1
U2.Czmot	0.21	0.04	5.12	# slope3 – slope2
U3.CZmot	-0.18	0.04	-4.37	# slope4 – slope3

Null deviance: 209.4 on 45 degrees of freedom
 Residual deviance: 46.7 on 38 degrees of freedom
 AIC: 149.27
Table A3.6 Model output for the Czech Republic

Appendix A4.1

Distance travelled by age class and travel mode

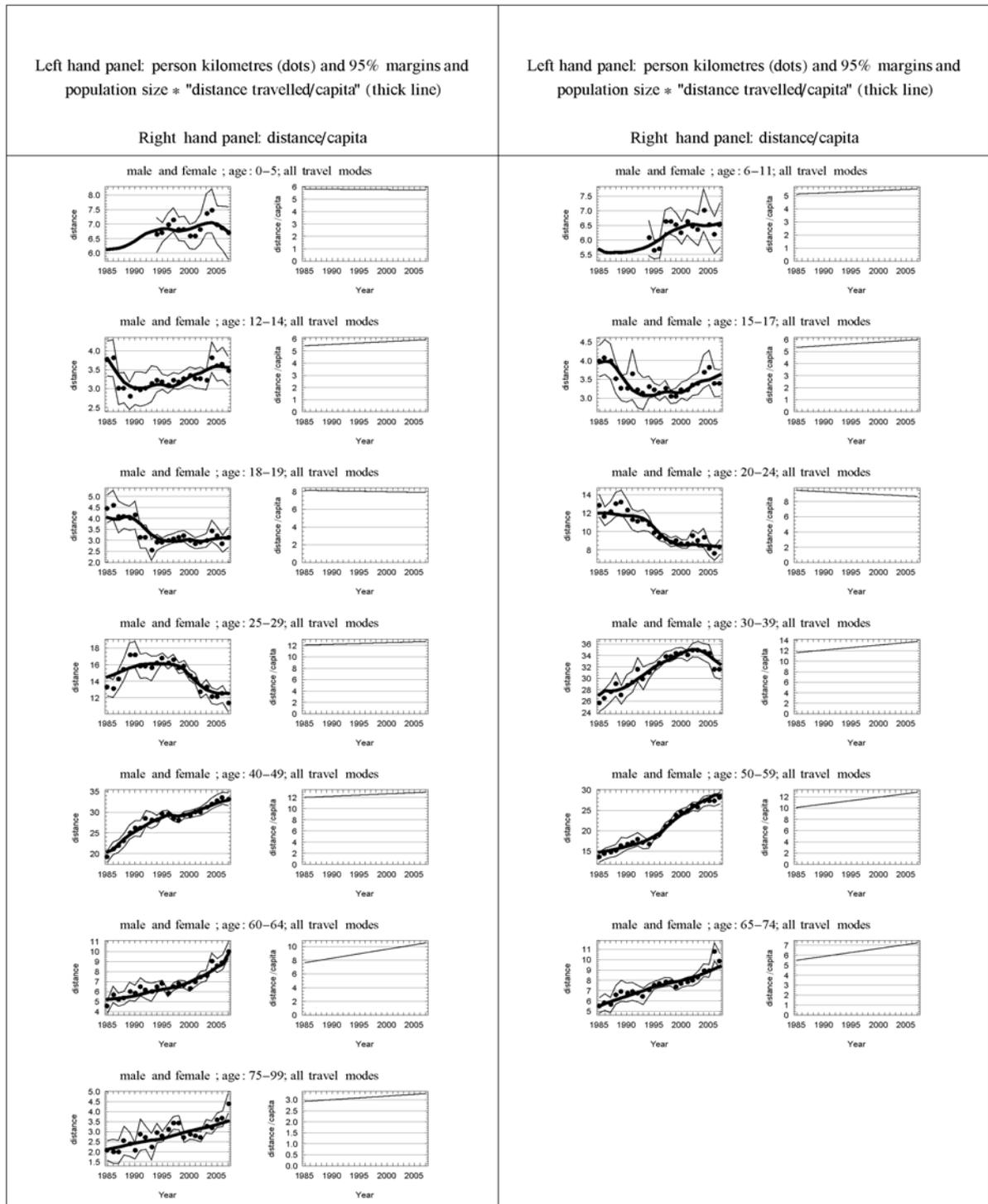


Figure A4.1.1 Annual distance travelled by all male or female inhabitants, by age group, both data and mode results. For each age group the left hand panel shows: total distance travelled (10^9 km, dots) and 95% confidence intervals for the estimate. The thick line represents the prediction by model (4.1). Each right hand panel shows per capita distance travelled (10^9 km) per million inhabitants according to the model. As the Dutch MON survey only started to collect

mobility data for road users younger than 12 years from 1994 onwards, the observations for 1985-1993 are (partly) missing for some of the graphs.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

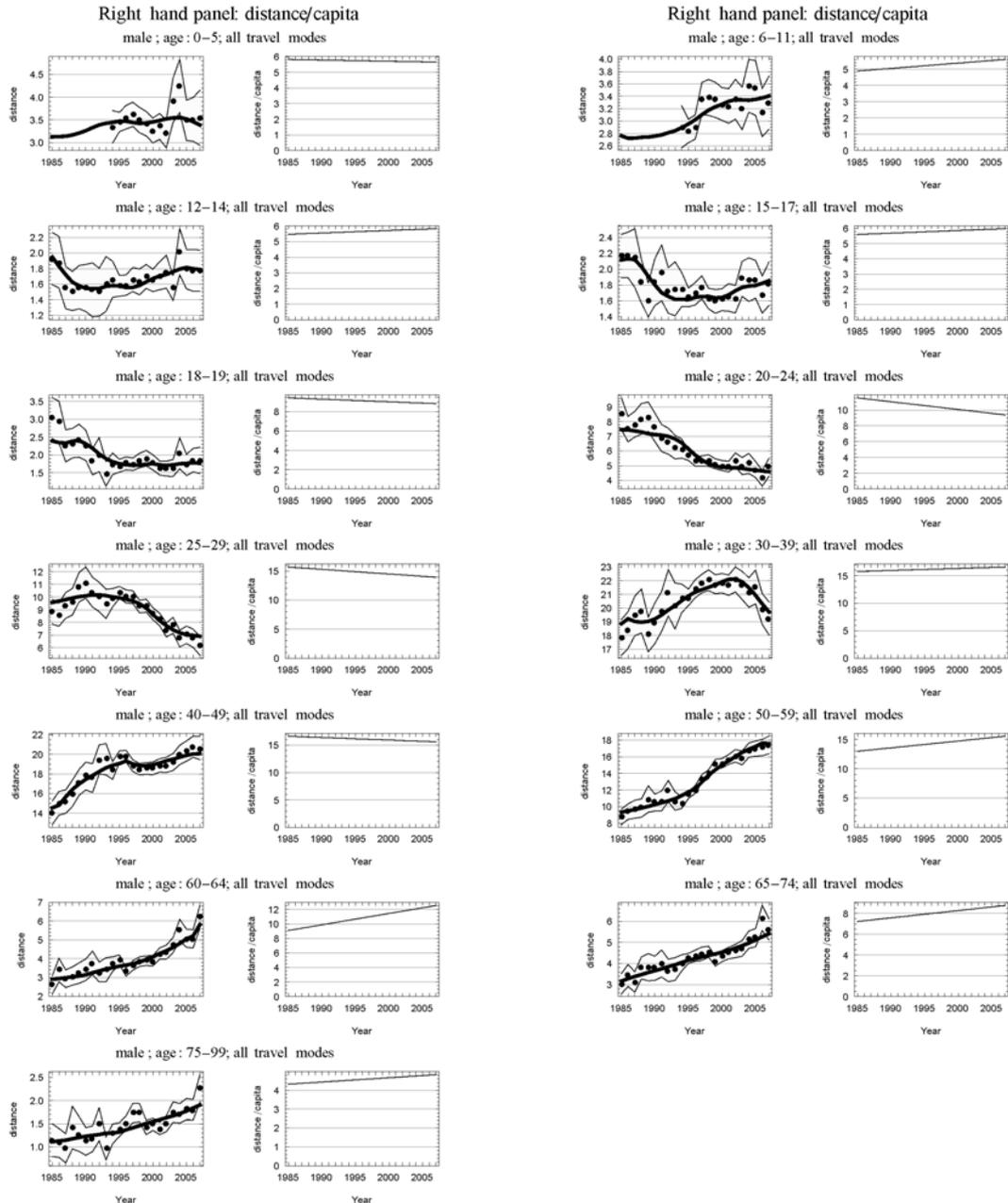


Figure A4.1.2 Annual distance travelled by all male inhabitants, by age group, both data and mode results. For further explanation: see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

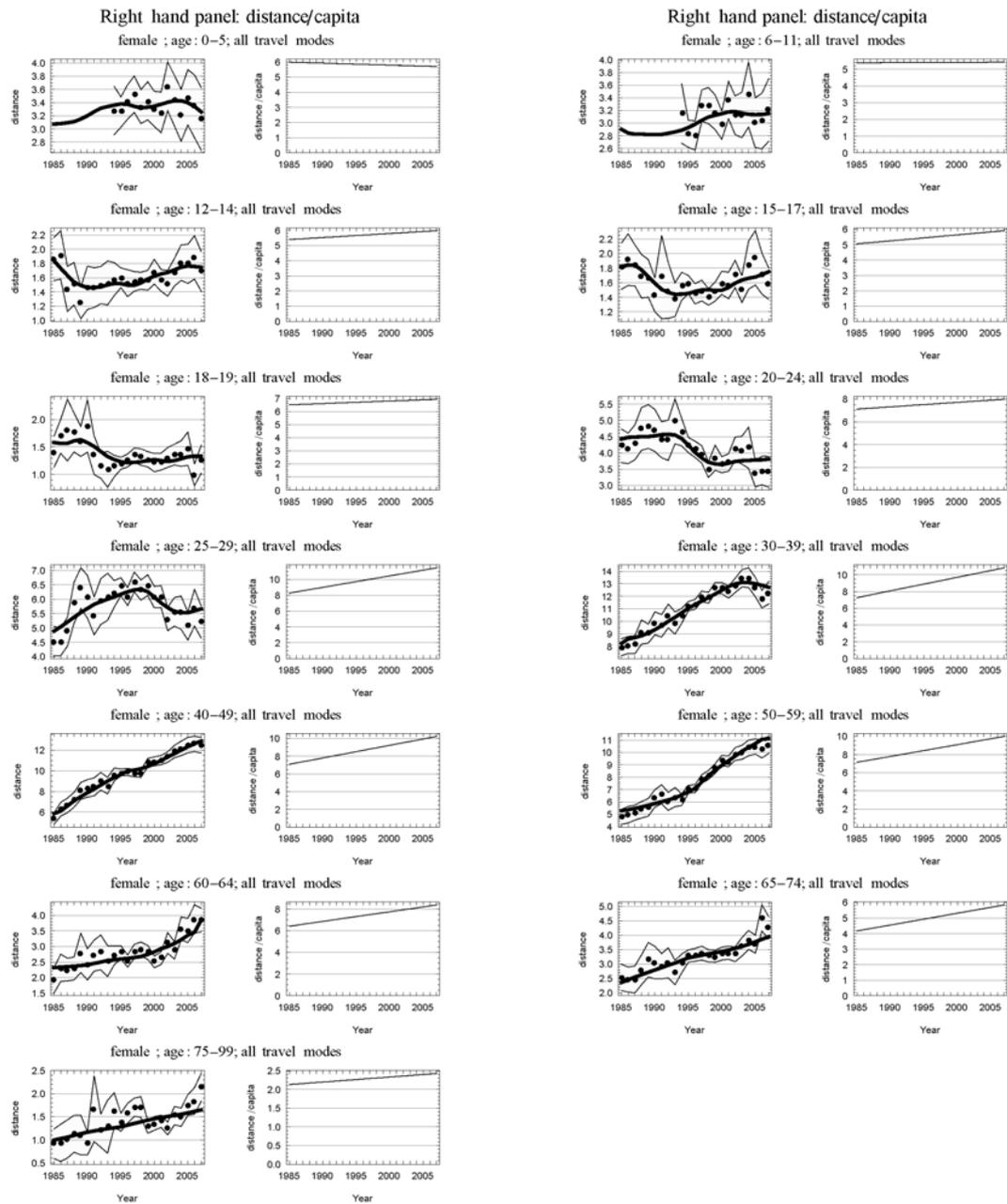


Figure A4.1.3 Annual distance travelled by all female inhabitants, by age group, both data and mode results. For further explanation: see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

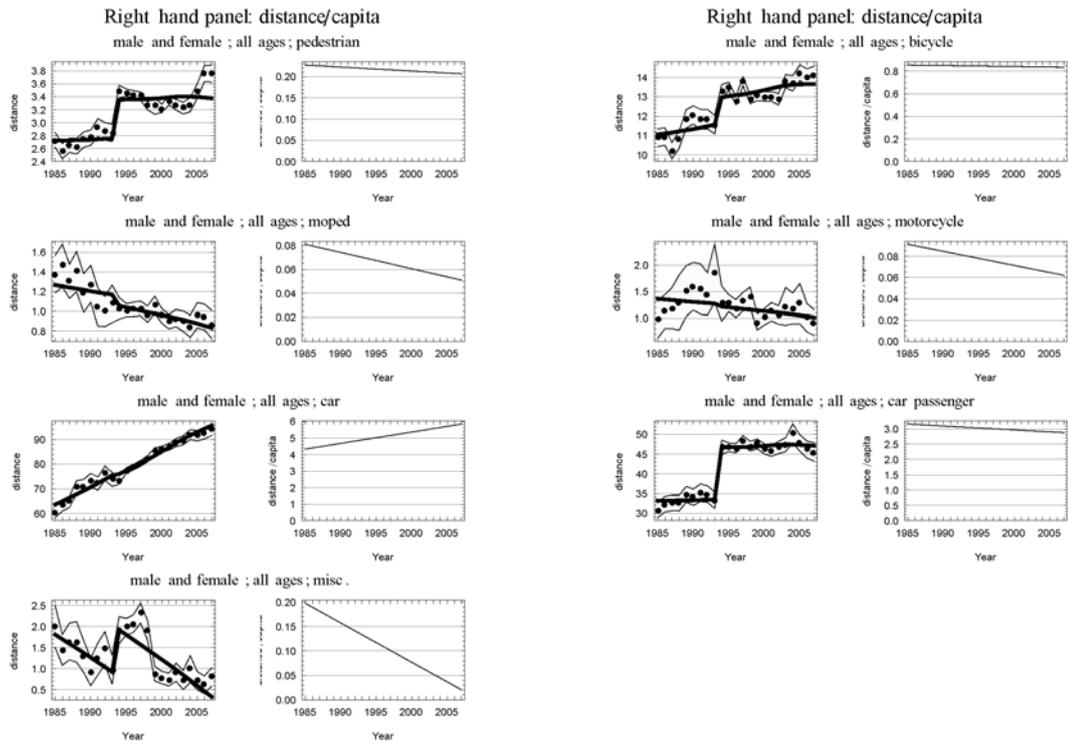


Figure A4.1.4 Annual distance travelled by all inhabitants for all ages by traffic mode. For further explanation see Figure A4.1.1

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

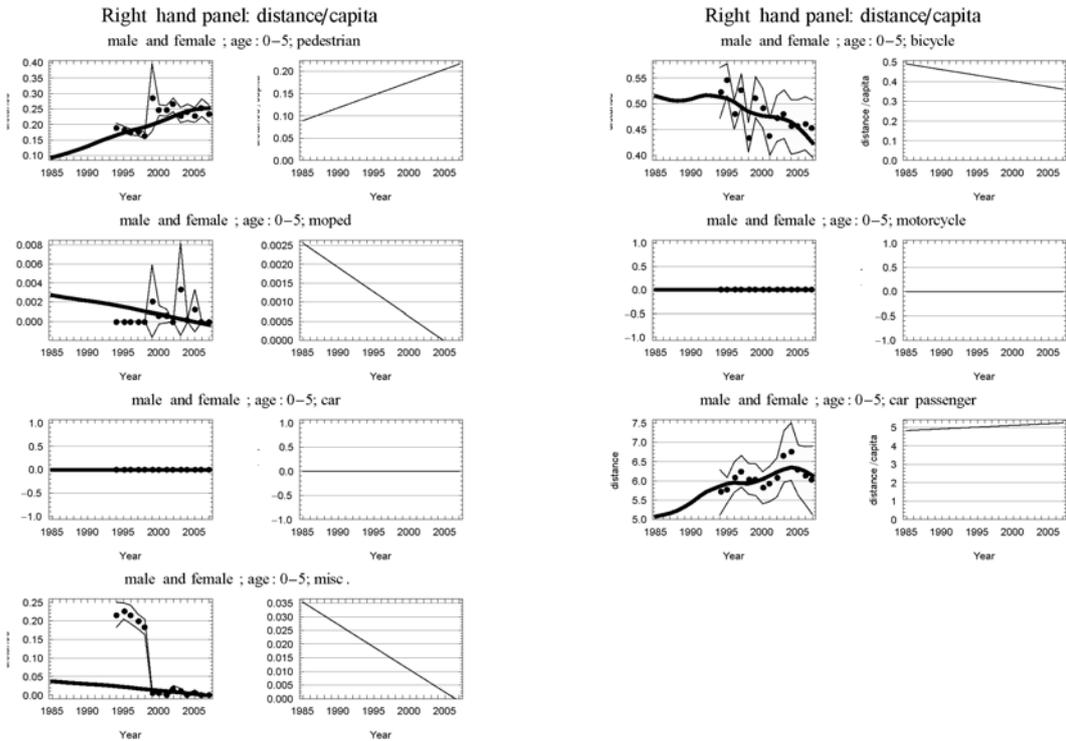


Figure A4.1.5 Annual distance travelled by all male or female inhabitants, 0-5 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

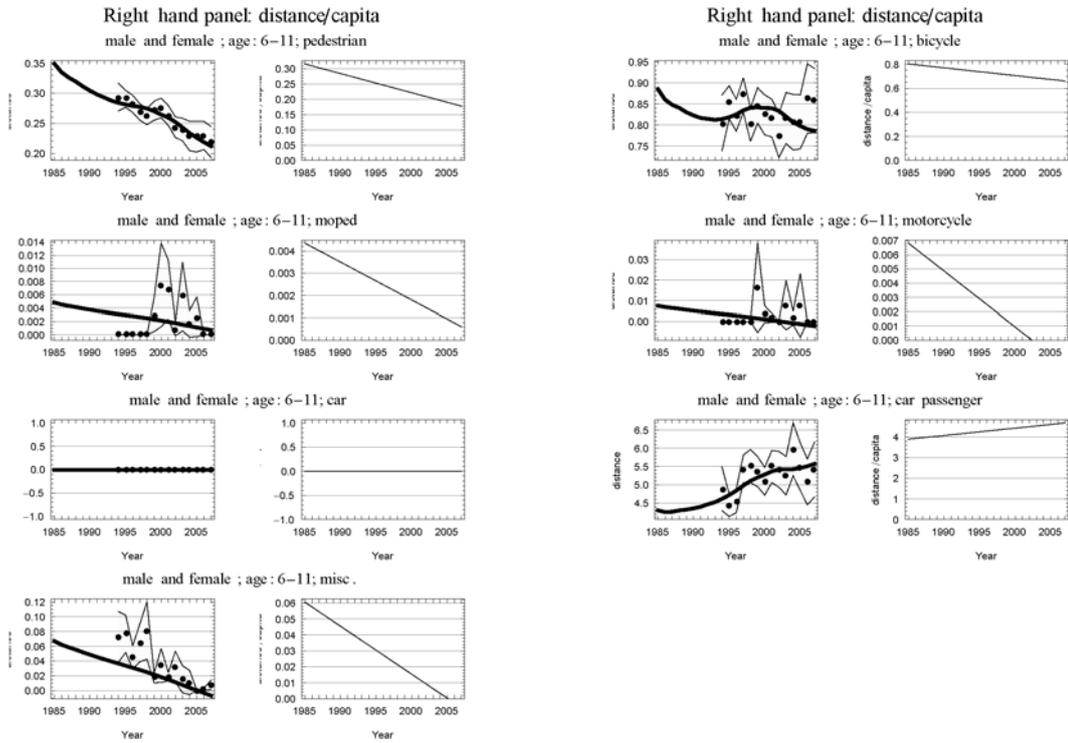


Figure A4.1.6 Annual distance travelled by all male or female inhabitants, 6-11 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

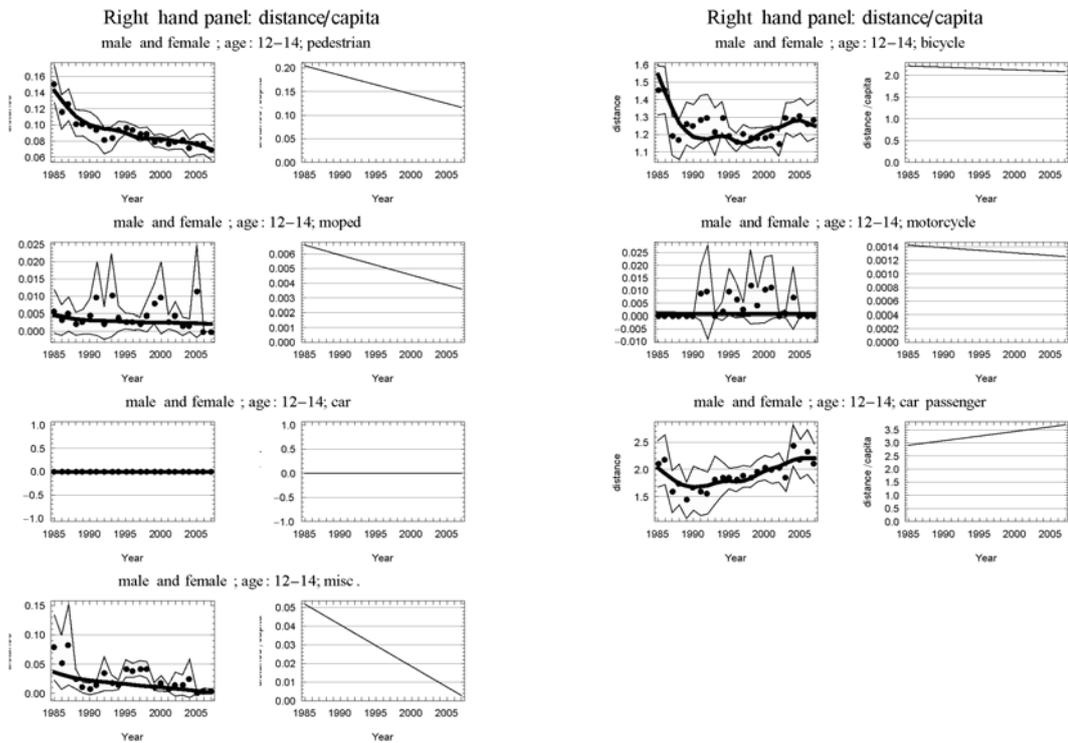


Figure A4.1.7 Annual distance travelled by all male or female inhabitants, 12-14 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

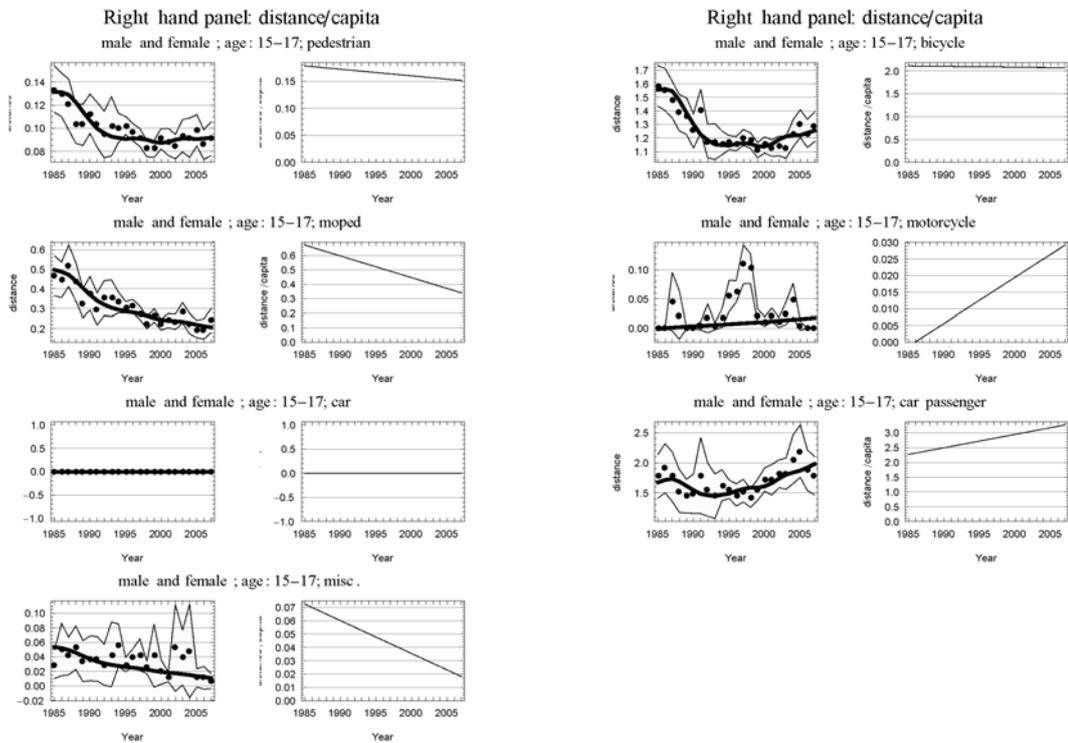


Figure A4.1.8 Annual distance travelled by all male or female inhabitants, 15-17 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

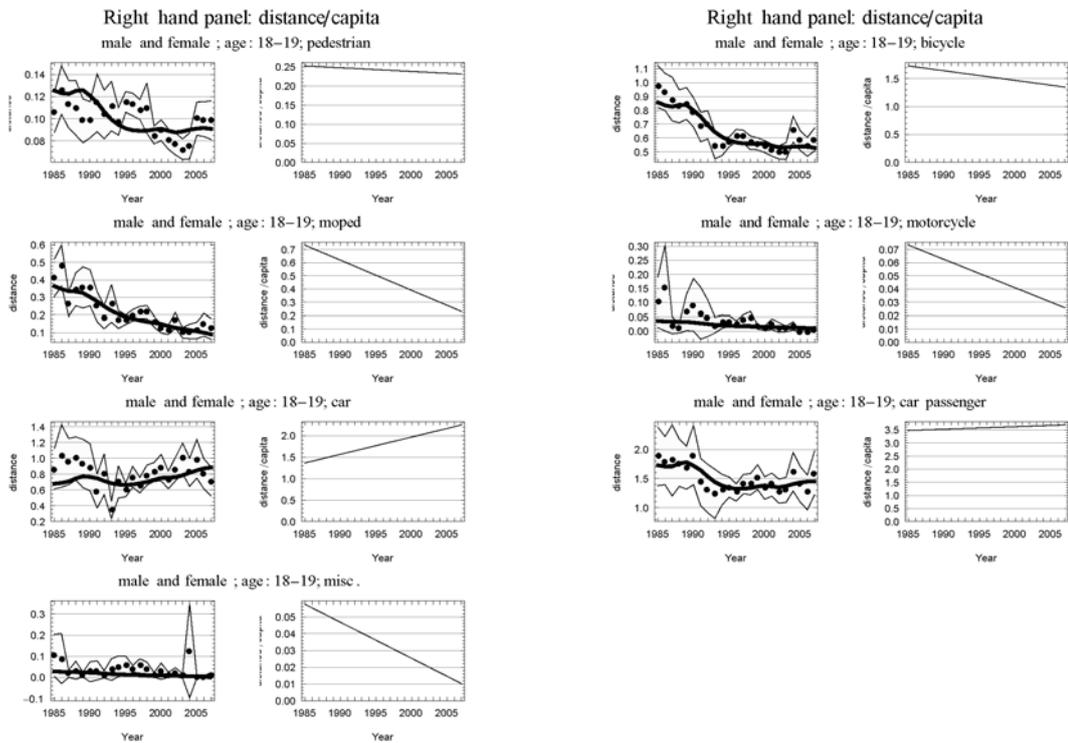


Figure A4.1.9 Annual distance travelled by all male or female inhabitants, 18-19 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

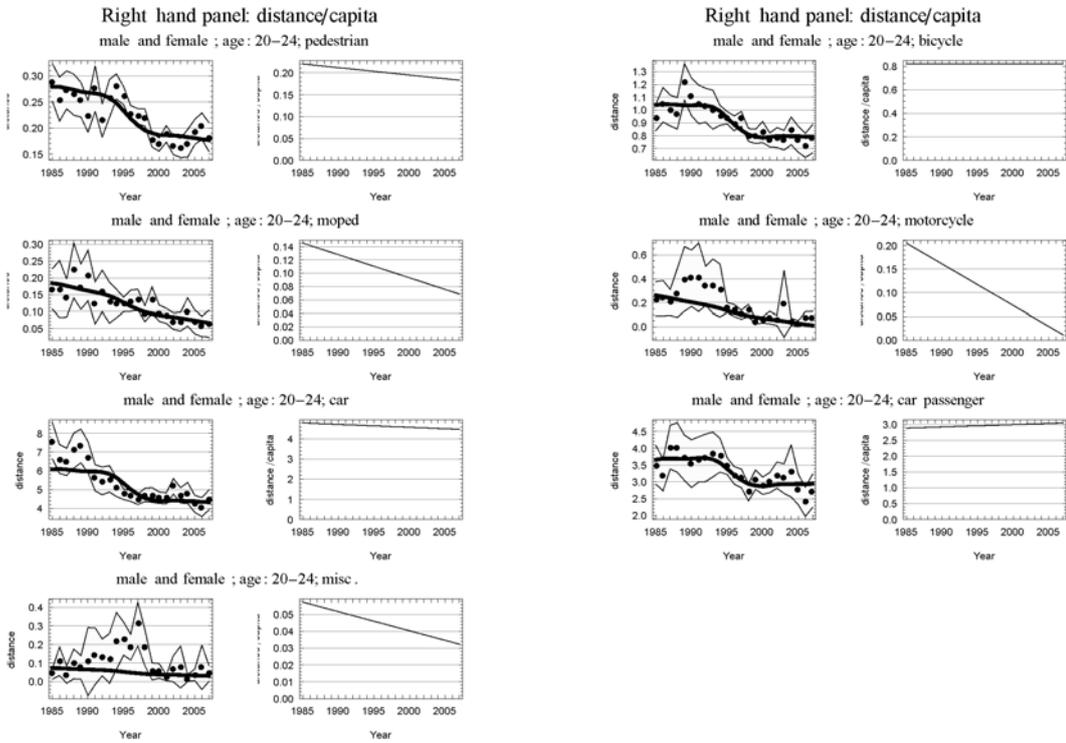


Figure A4.1.10 Annual distance travelled by all male or female inhabitants, 20-24 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

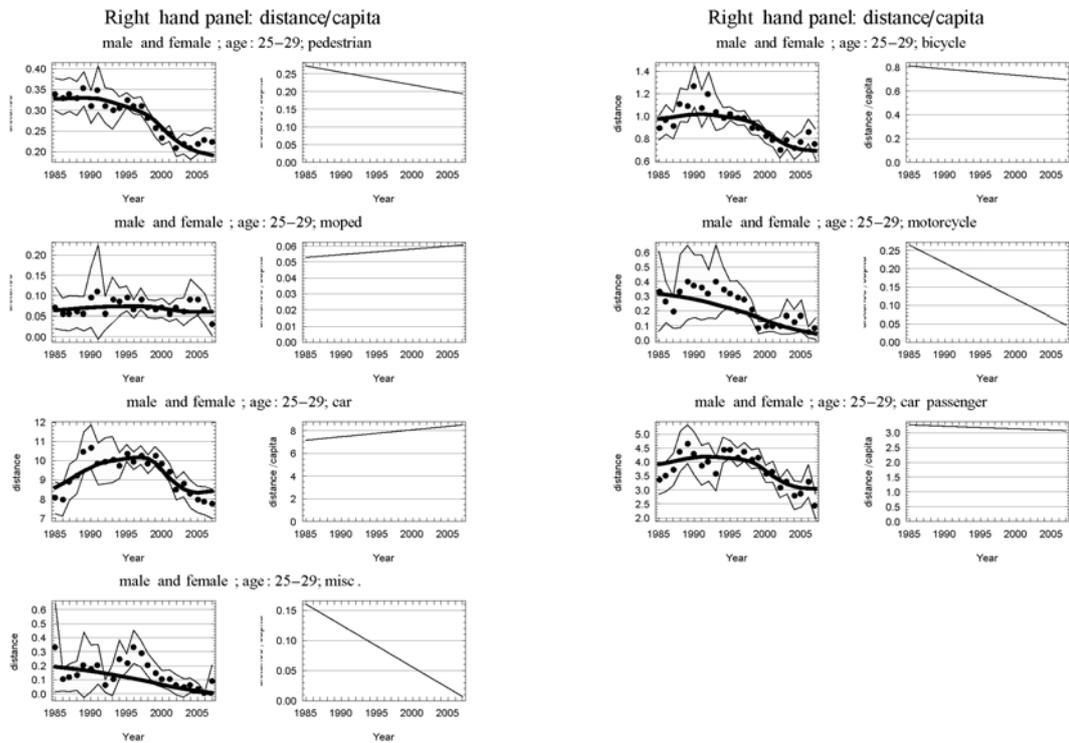


Figure A4.1.11 Annual distance travelled by all male or female inhabitants, 25-29 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

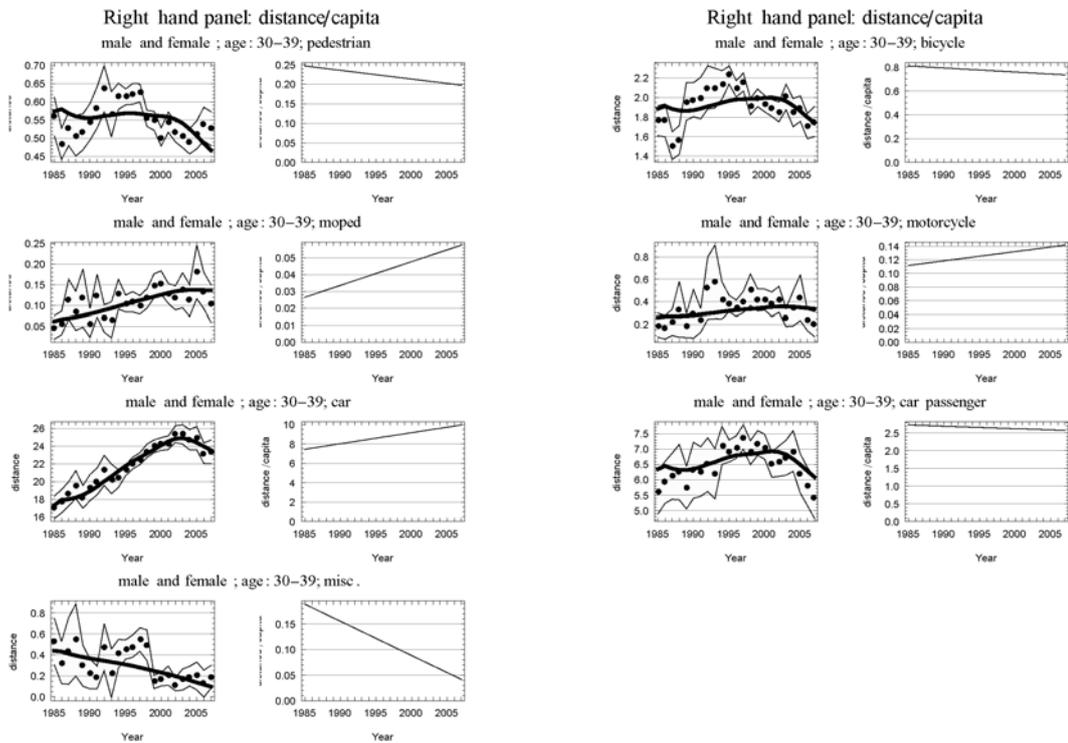


Figure A4.1.12 Annual distance travelled by all male or female inhabitants, 30-39 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

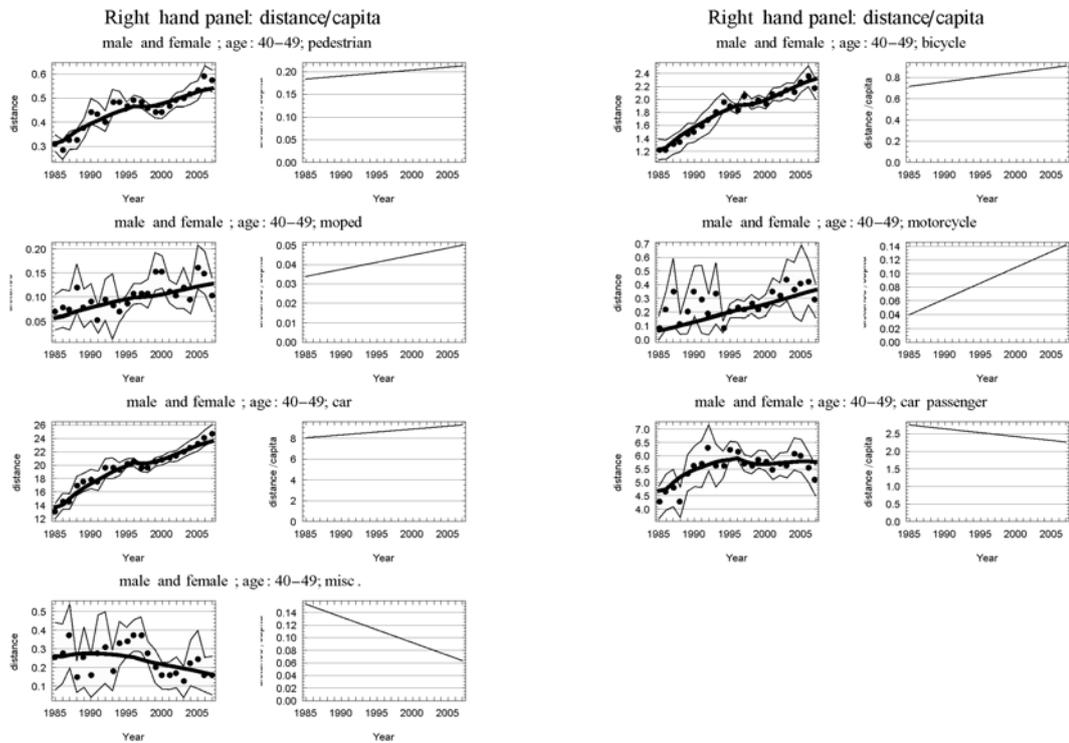


Figure A4.1.13 Annual distance travelled by all male or female inhabitants, 40-49 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

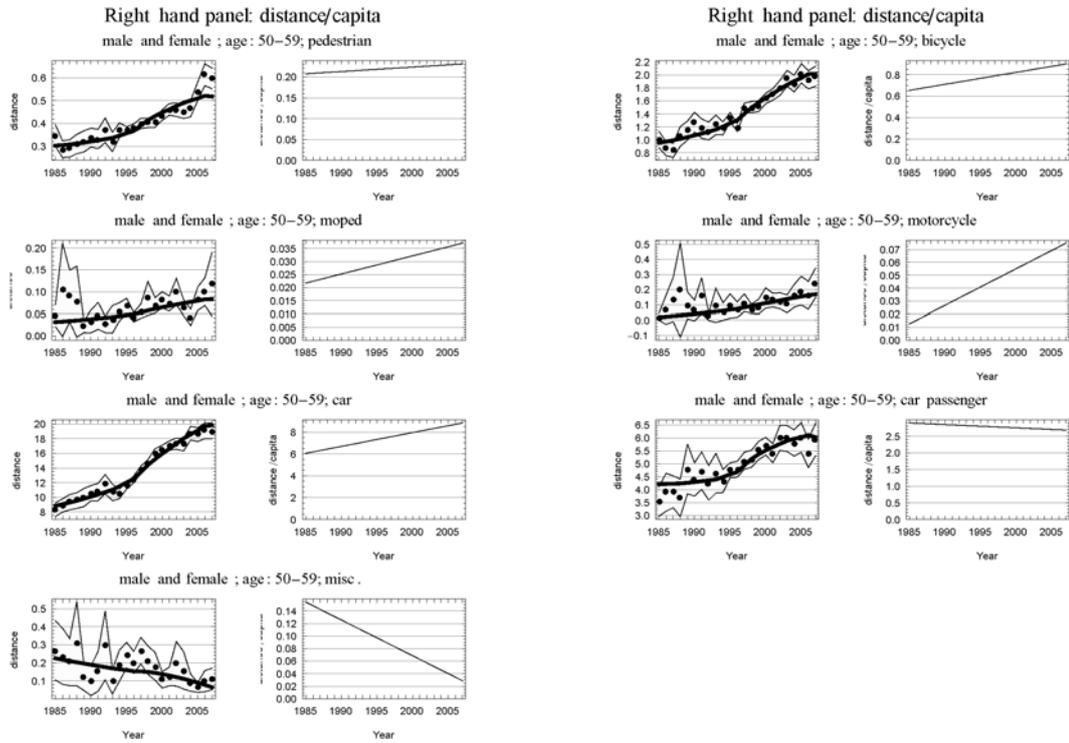


Figure A4.1.14 Annual distance travelled by all male or female inhabitants, 50-59 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

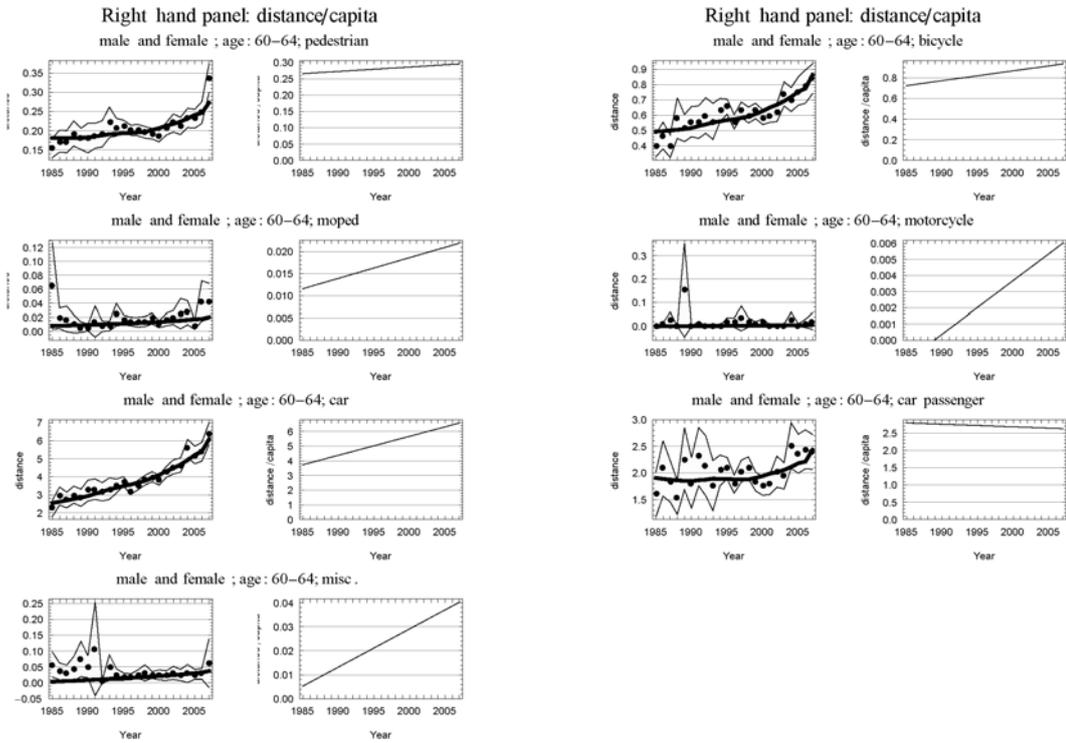


Figure A4.1.15 Annual distance travelled by all male or female inhabitants, 60-64 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

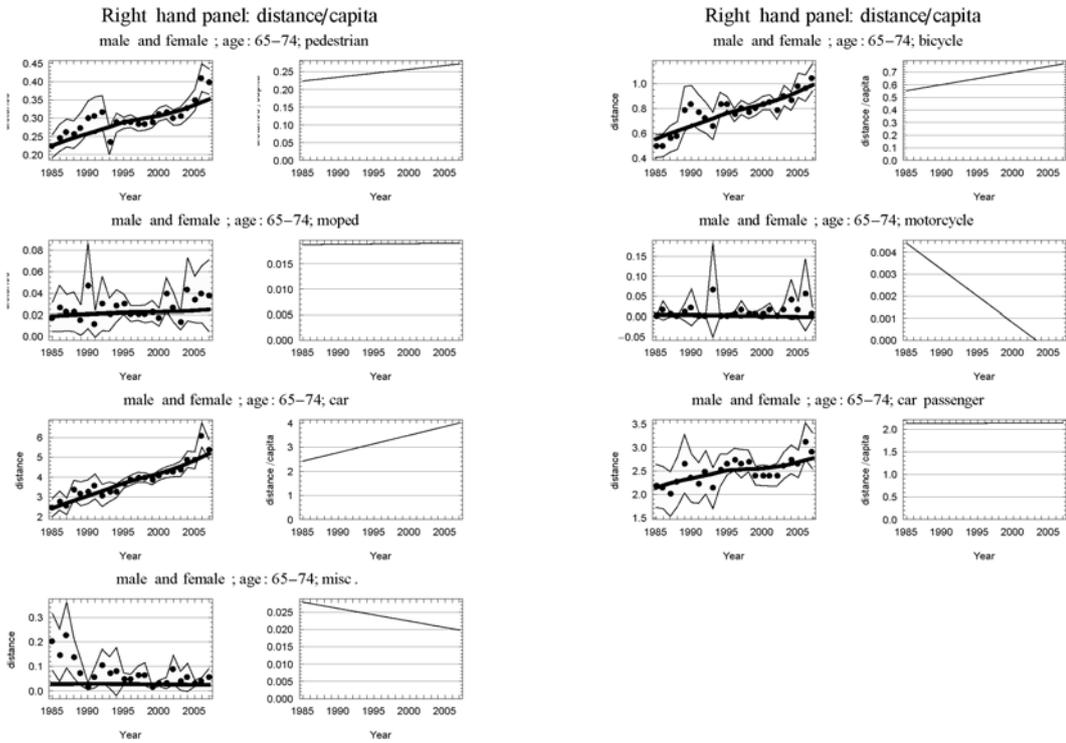


Figure A4.1.16 Annual distance travelled by all male or female inhabitants, 65-74 year old, by traffic mode. For further explanation see Figure A4.1.1.

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

Left hand panel: person kilometres (dots) and 95% margins and population size * "distance travelled/capita" (thick line)

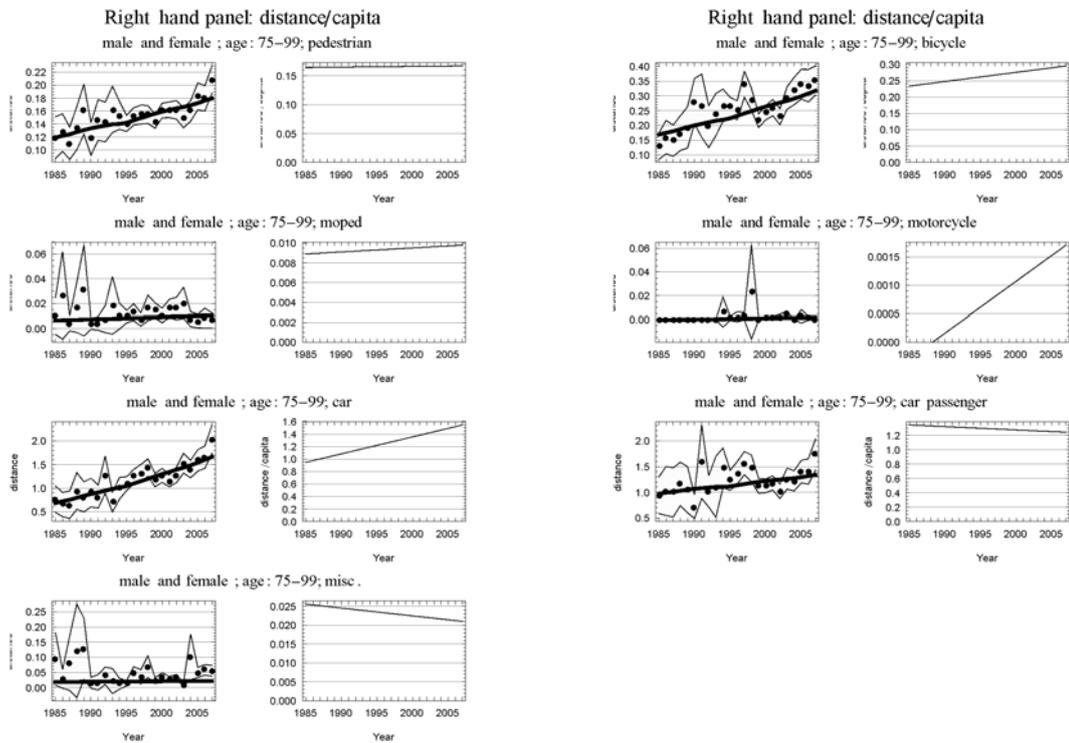


Figure A4.1.17 Annual distance travelled by all male or female inhabitants, 75+ year old, by traffic mode. For further explanation see Figure A4.1.1.

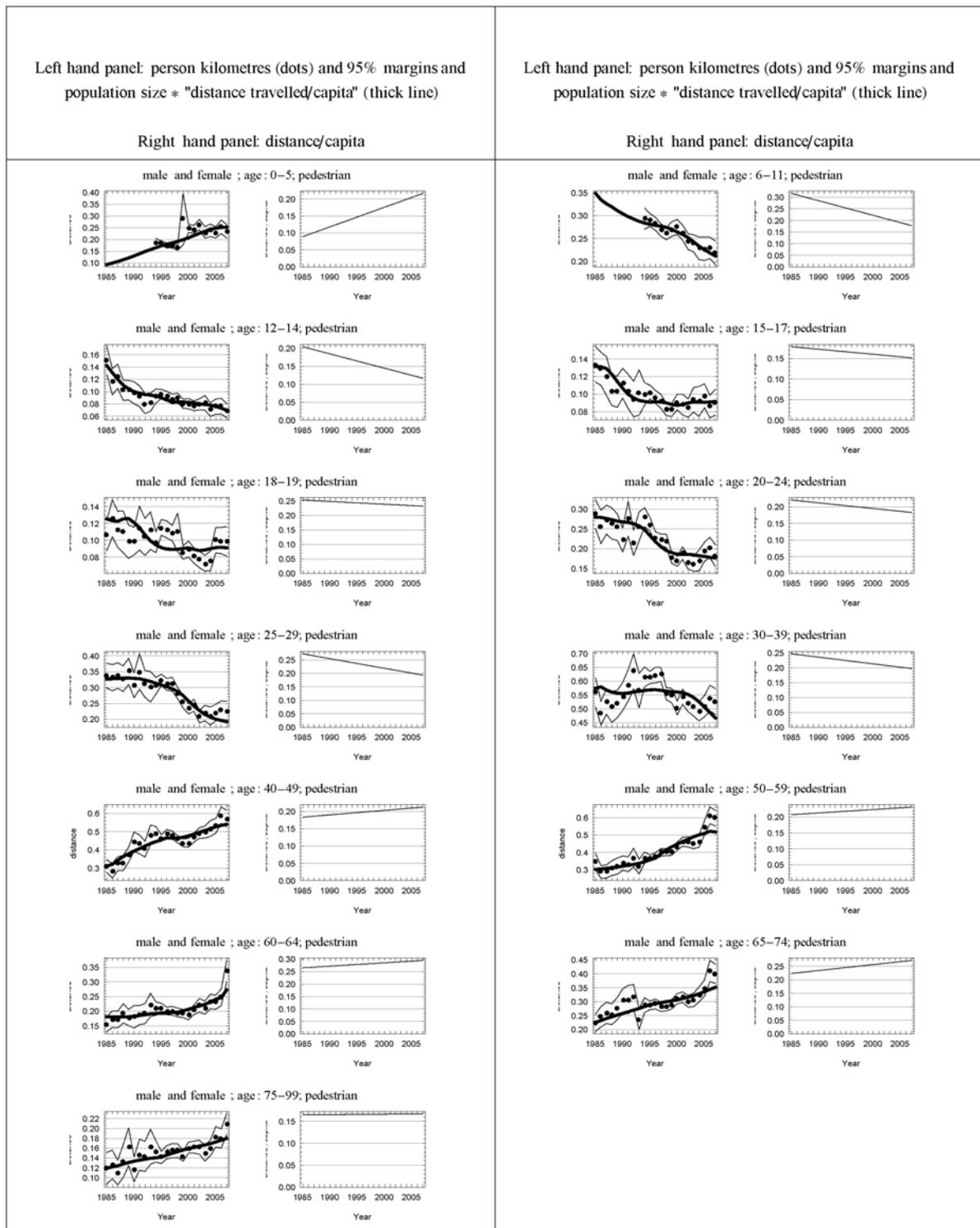


Figure A4.1.18 Annual pedestrian distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

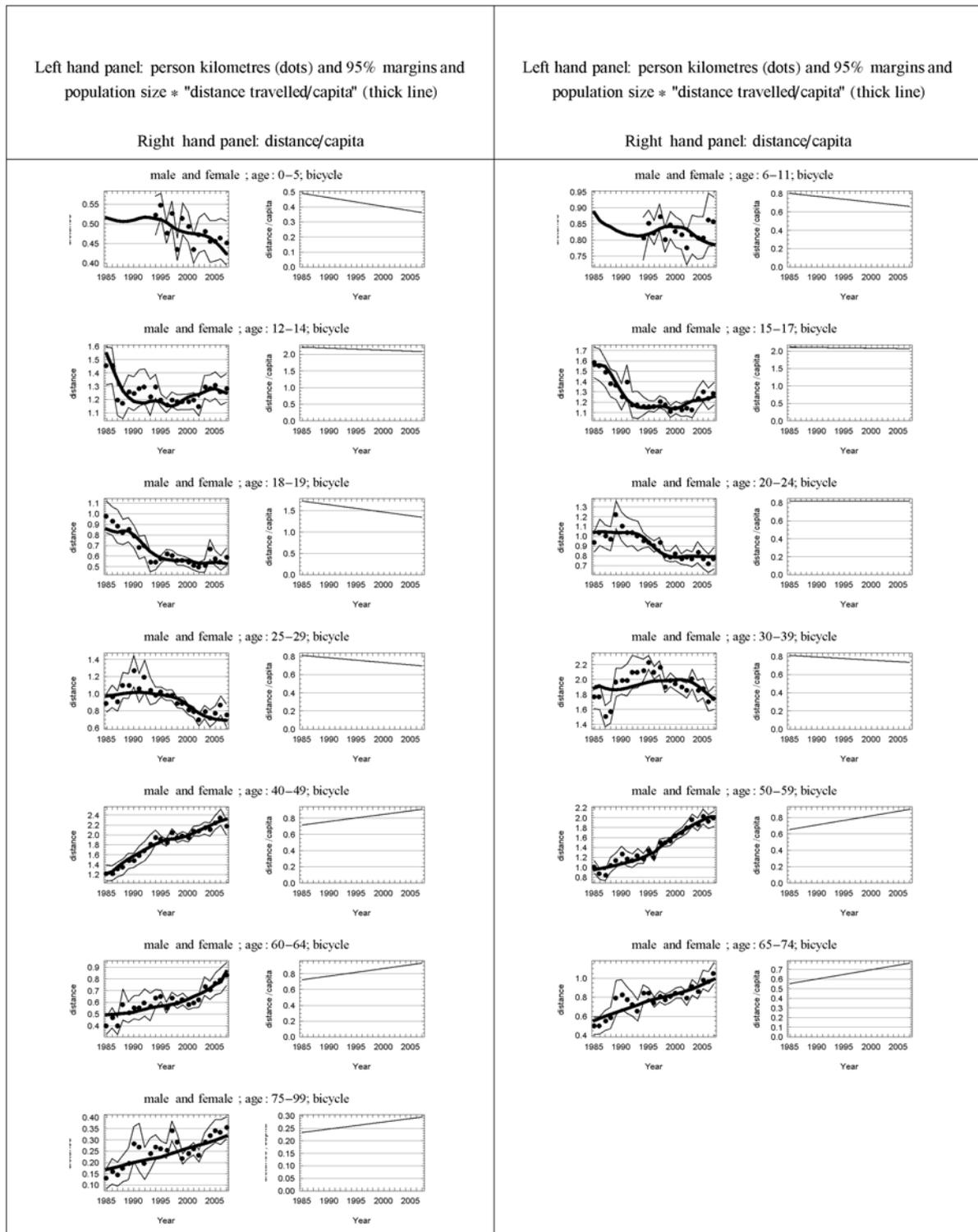


Figure A4.1.19 Annual bicycle distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

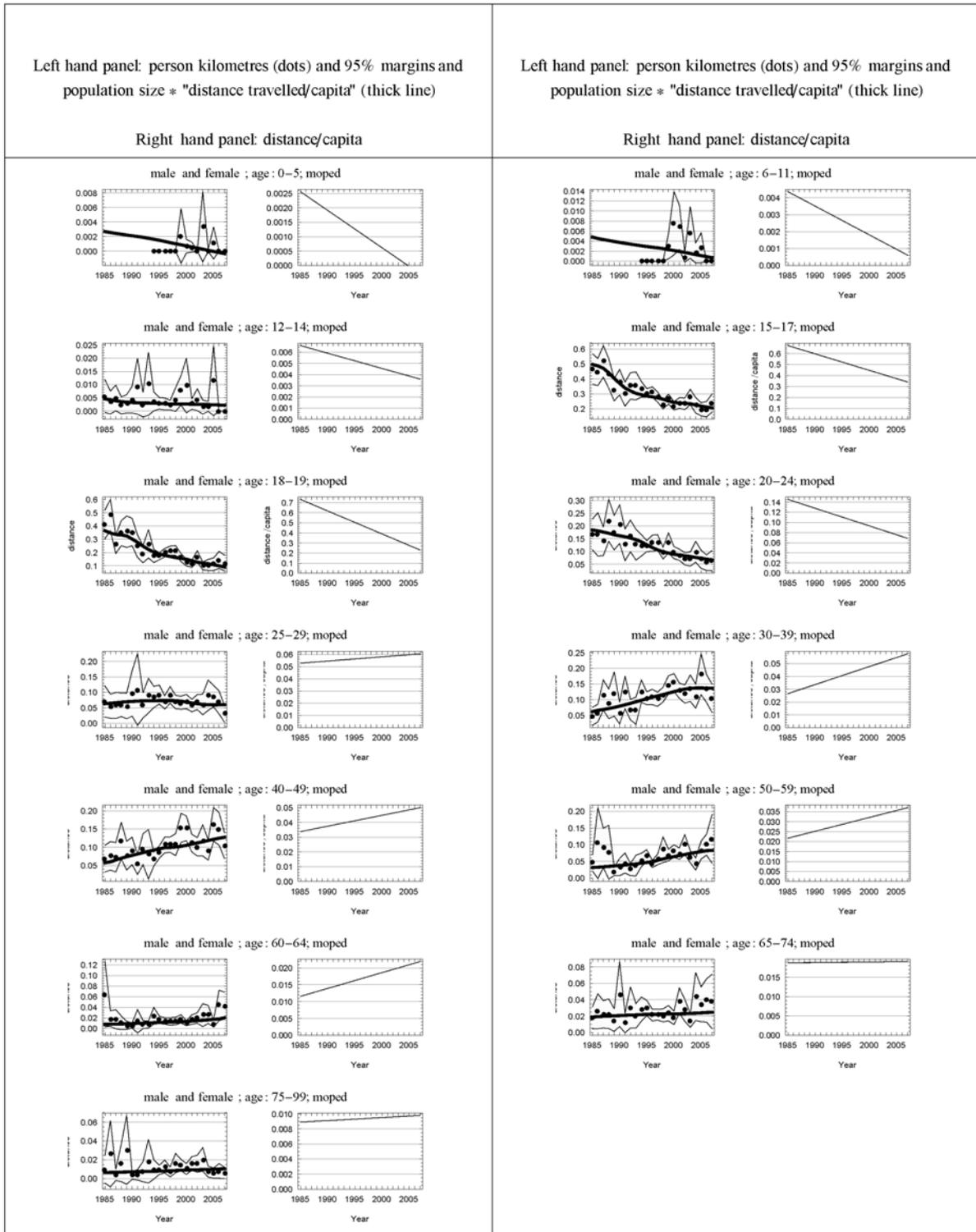


Figure A4.1.20 Annual moped distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

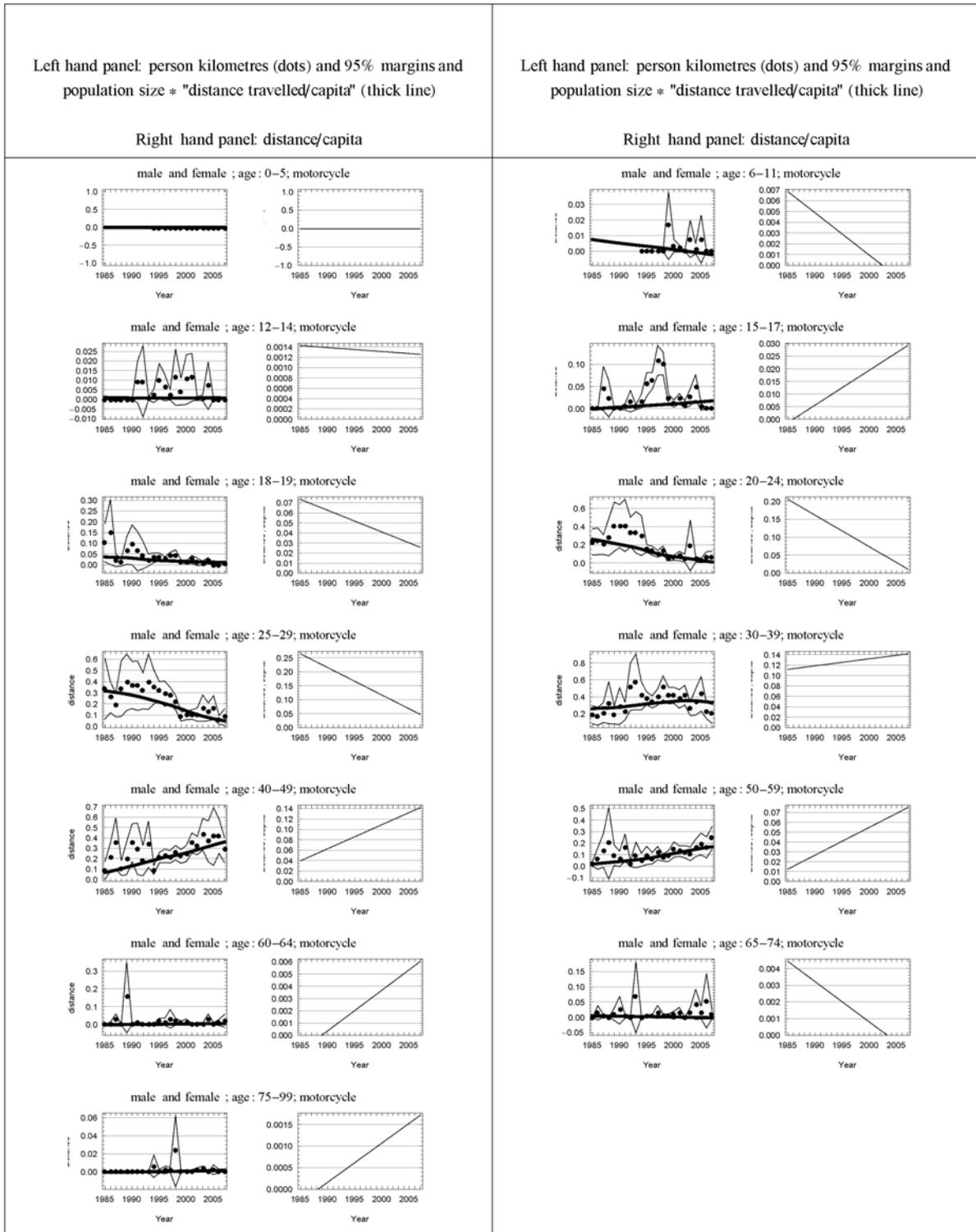


Figure A4.1.21 Annual motorcycle distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

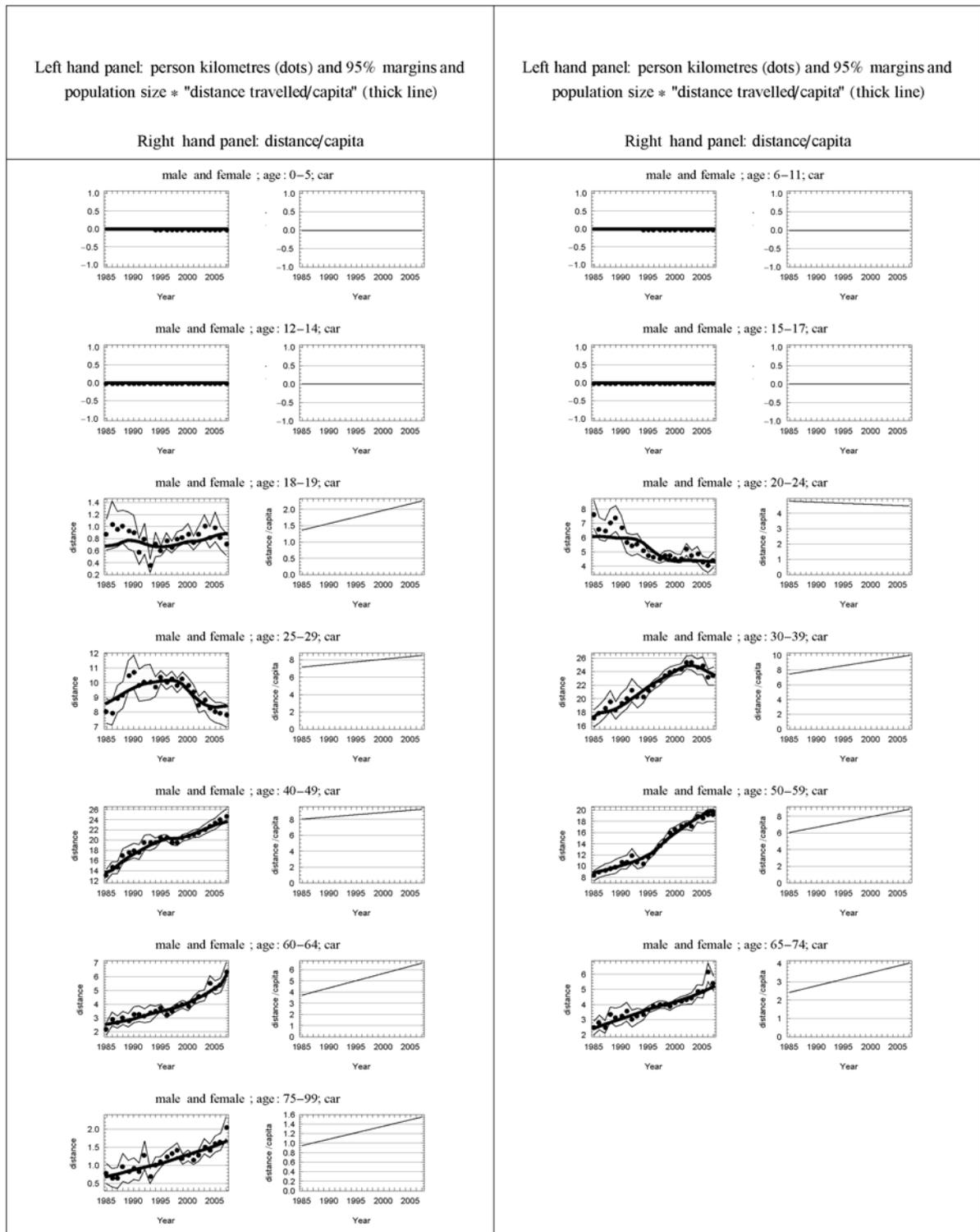


Figure A4.1.22 Annual passenger distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

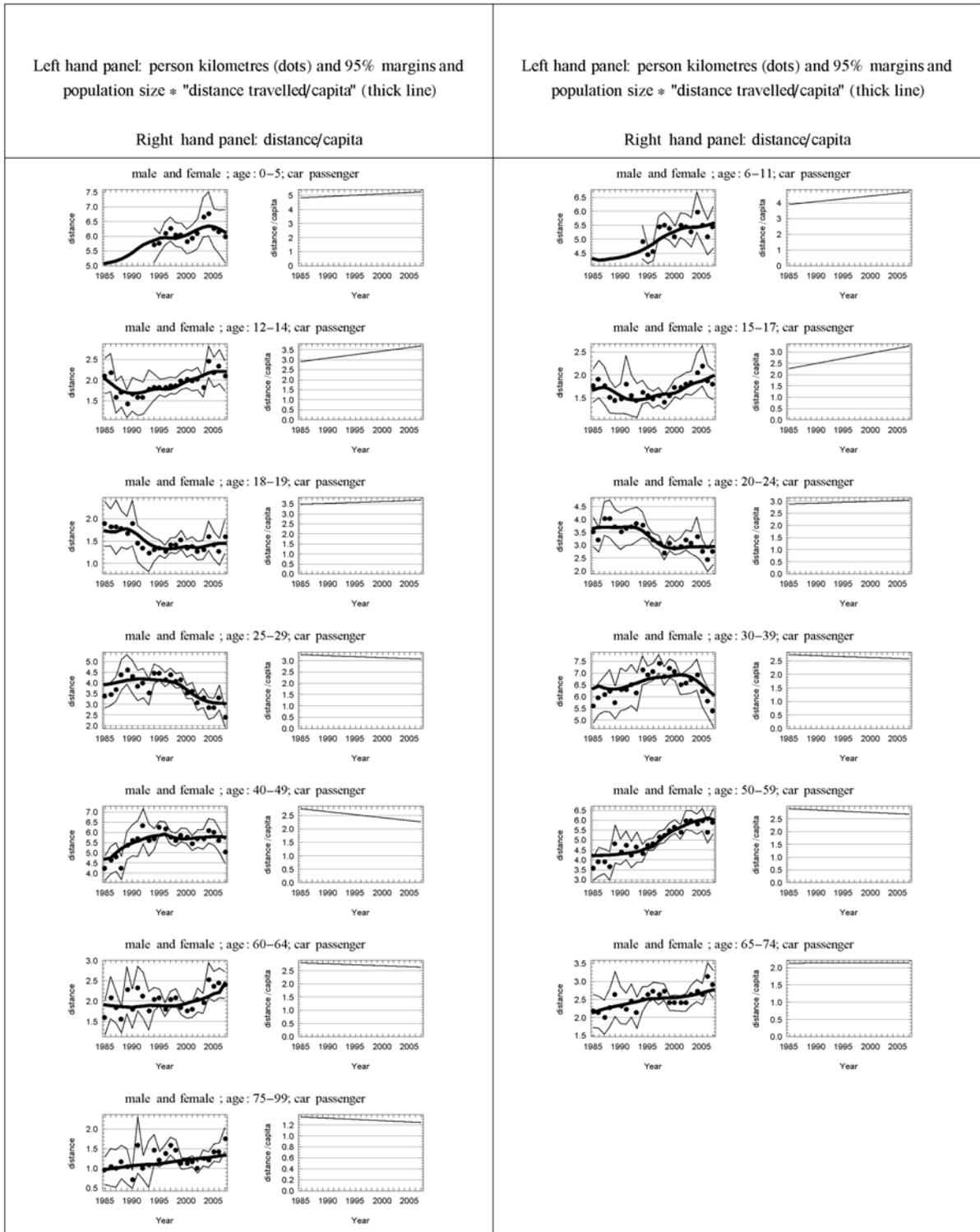


Figure A4.1.23 Annual passenger car distance travelled, by all male or female inhabitants by age. For further explanation see Figure A4.1.1.

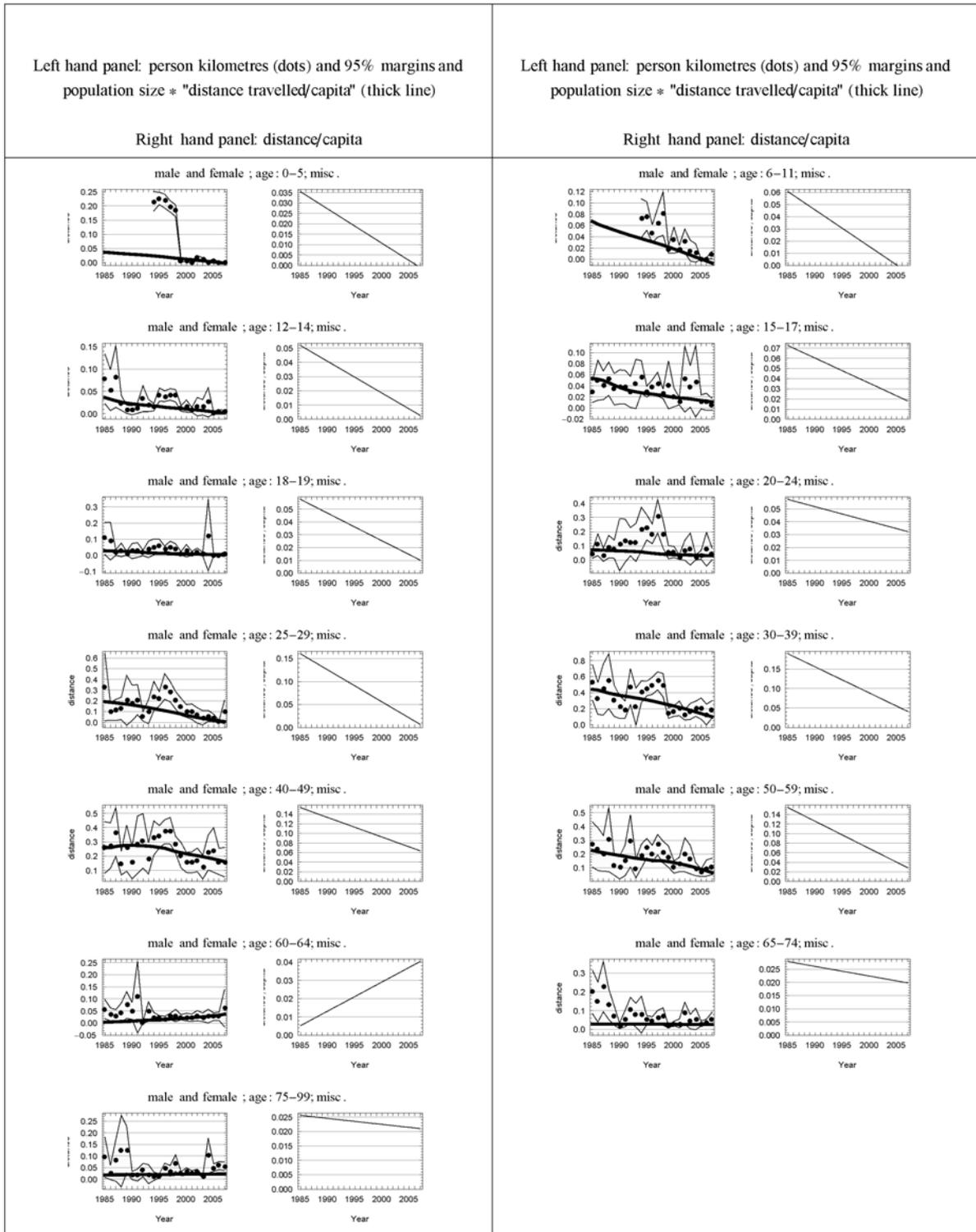


Figure A4.1.24 Annual distance travelled in miscellaneous vehicles, by all male or female inhabitants, by age. For further explanation see Figure A4.1.1.

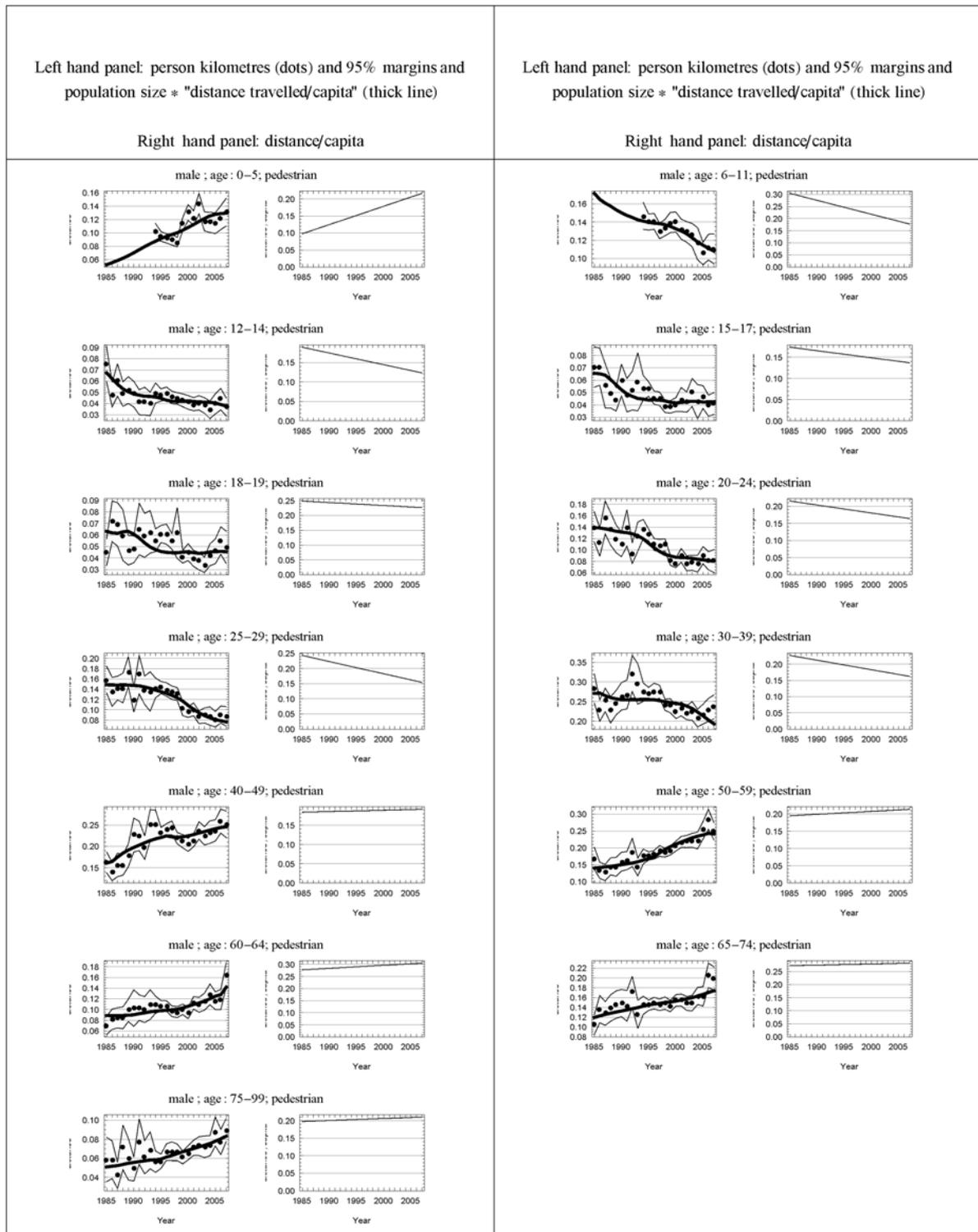


Figure A4.1.25 Annual pedestrian distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

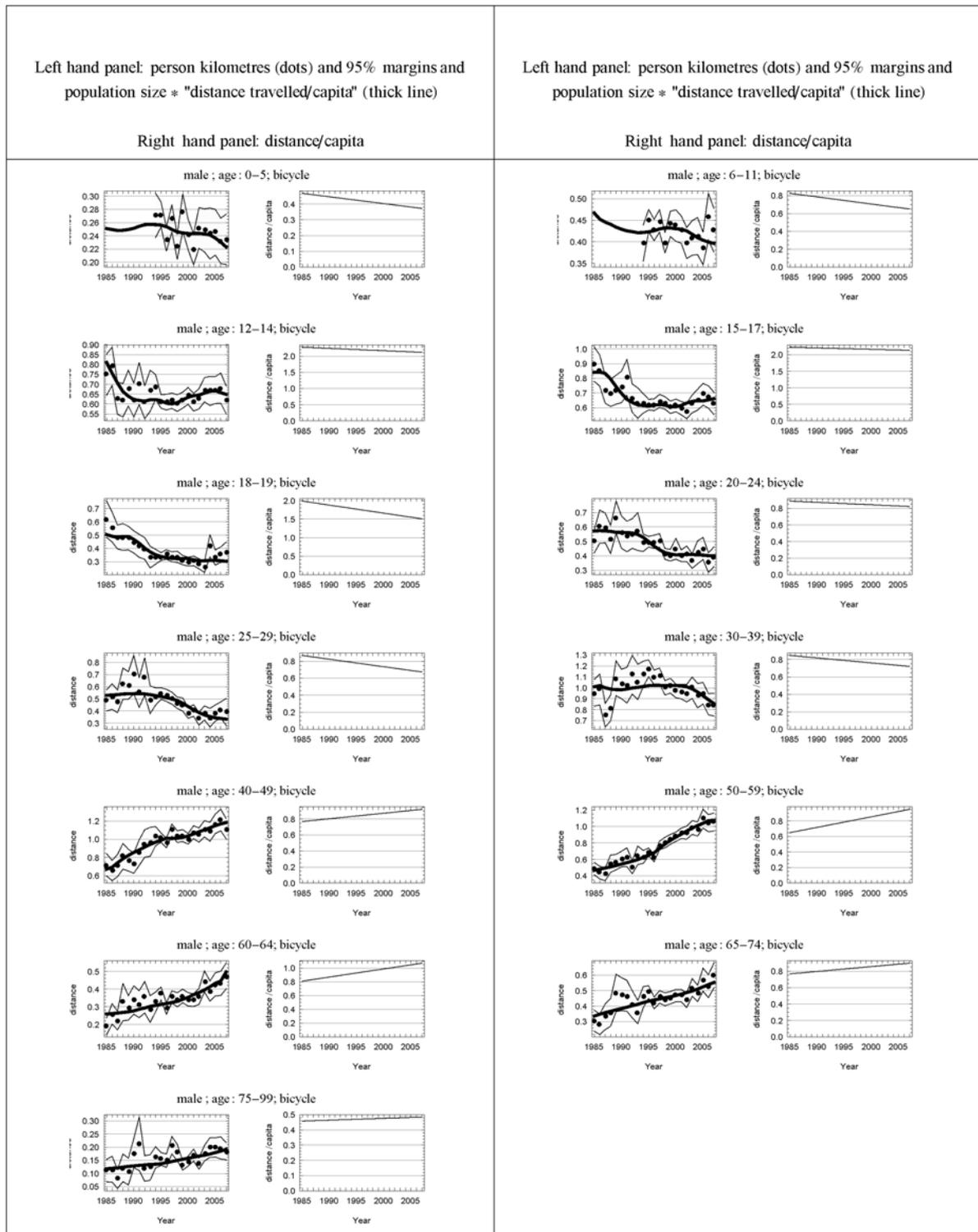


Figure A4.1.26 Annual bicycle distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

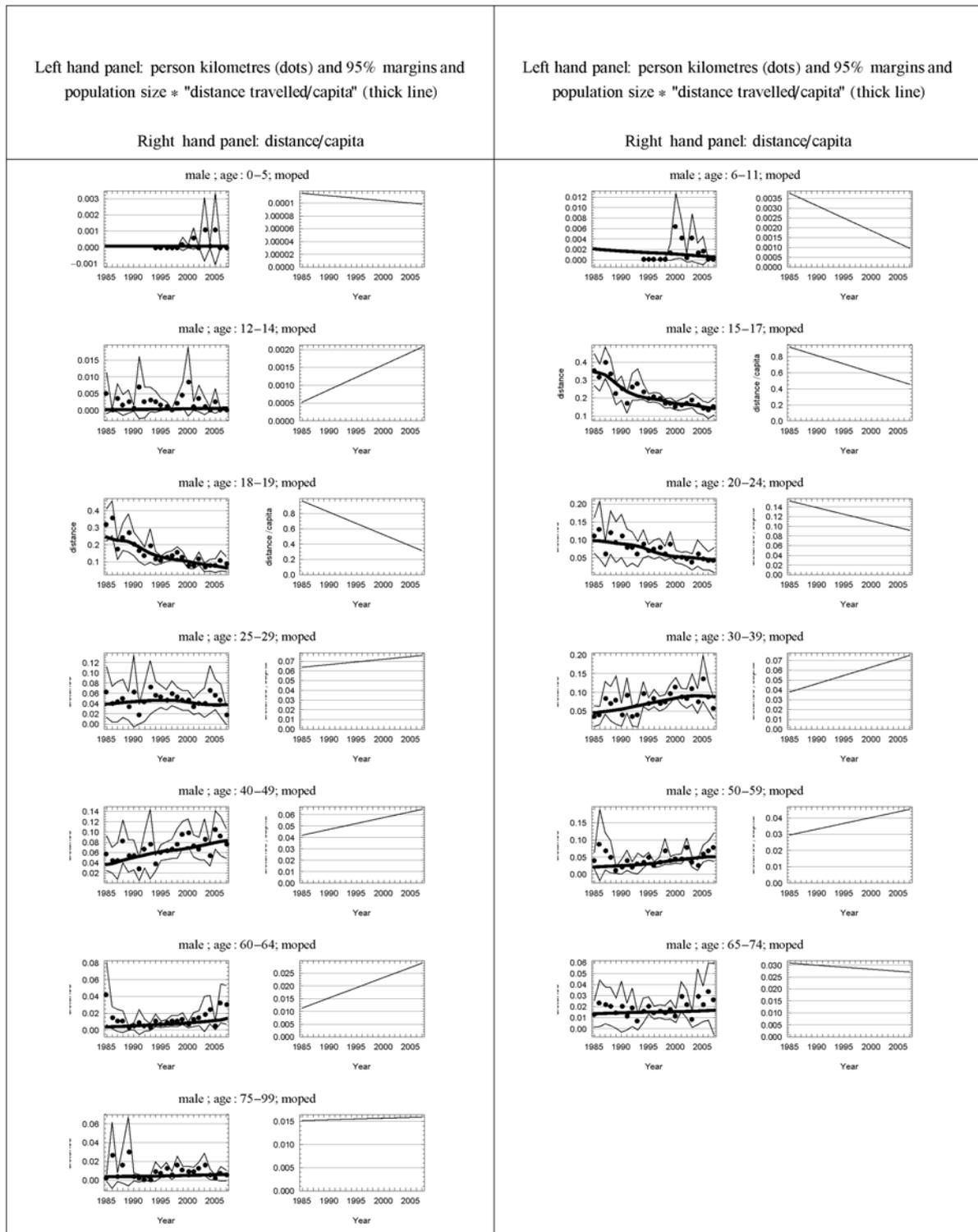


Figure A4.1.27 Annual moped distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

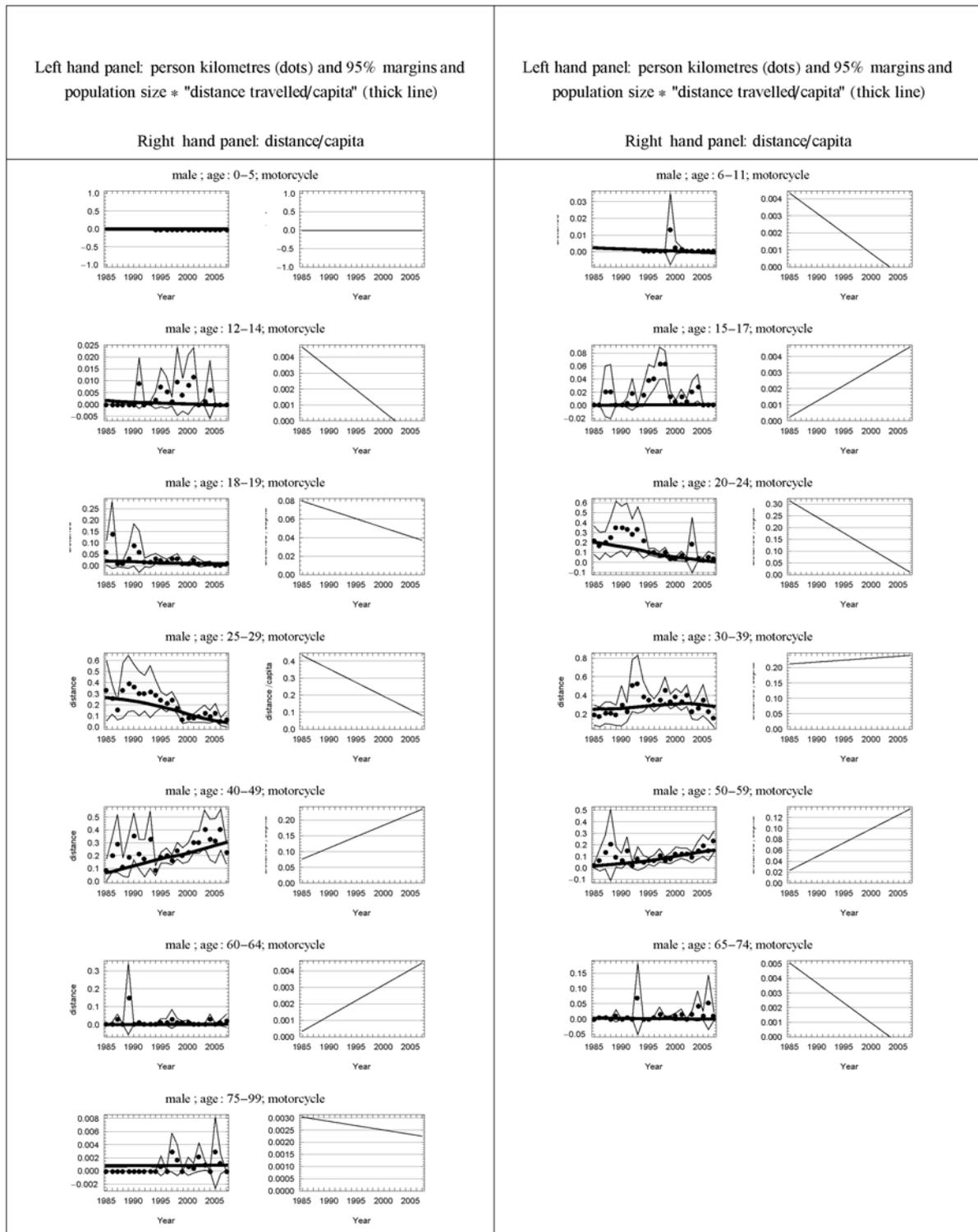


Figure A4.1.28 Annual motorcycle distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

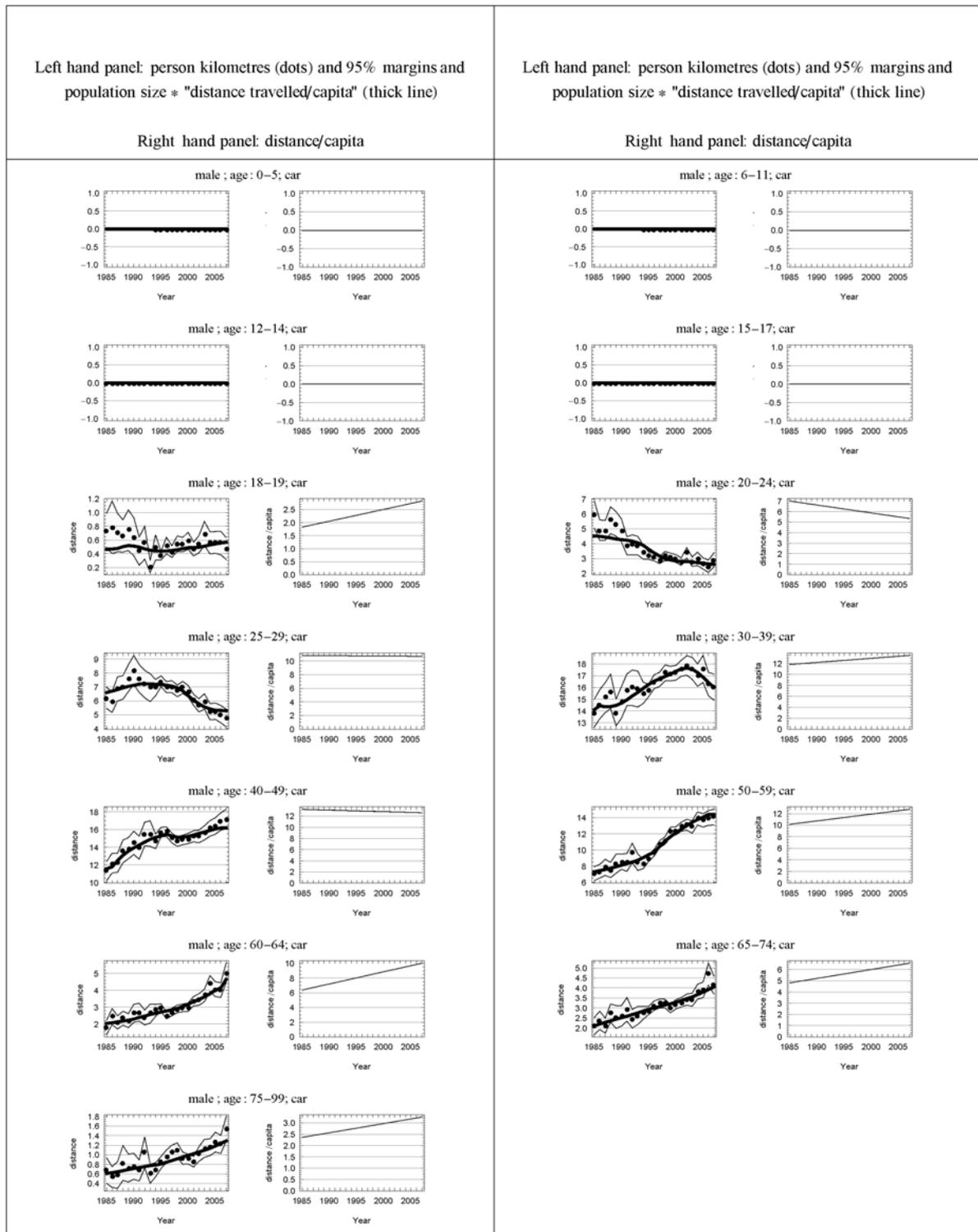


Figure A4.1.29 Annual passenger car driver distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

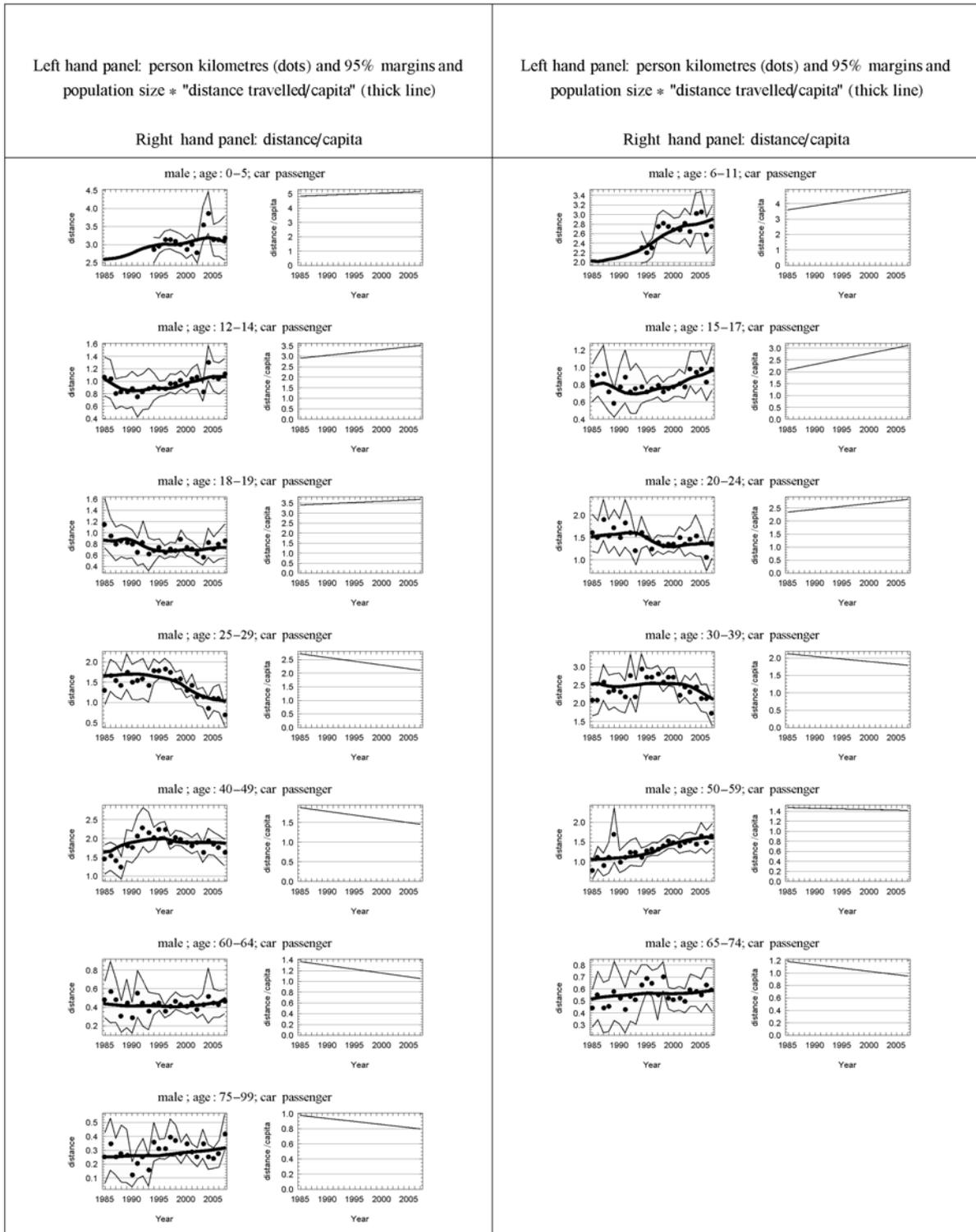


Figure A4.1.30 Annual car passenger distance travelled, by male inhabitants, by age. For further explanation see Figure A4.1.1.

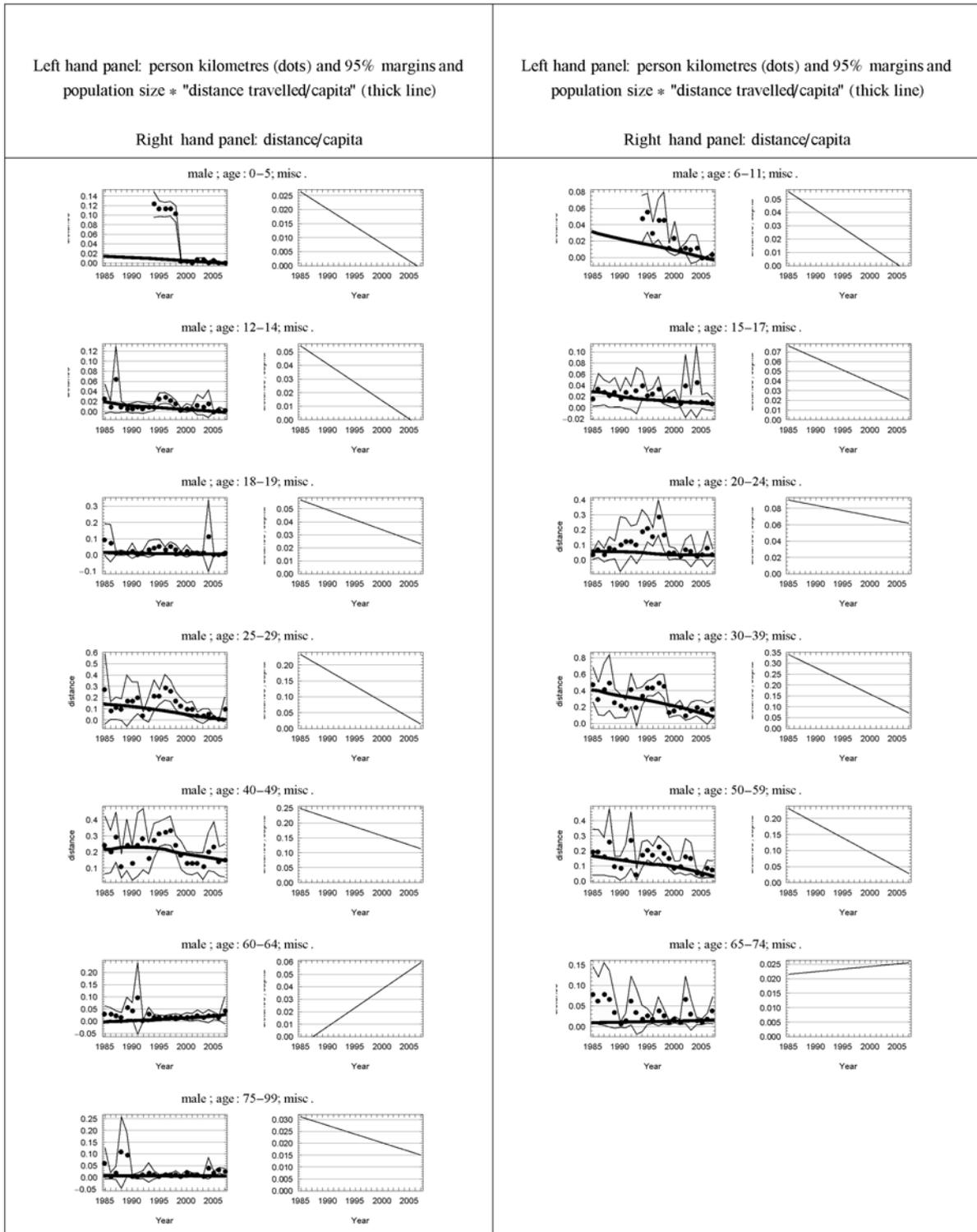


Figure A4.1.31 Annual distance travelled in miscellaneous vehicles, by male inhabitants, by age. For further explanation see Figure A4.1.1.

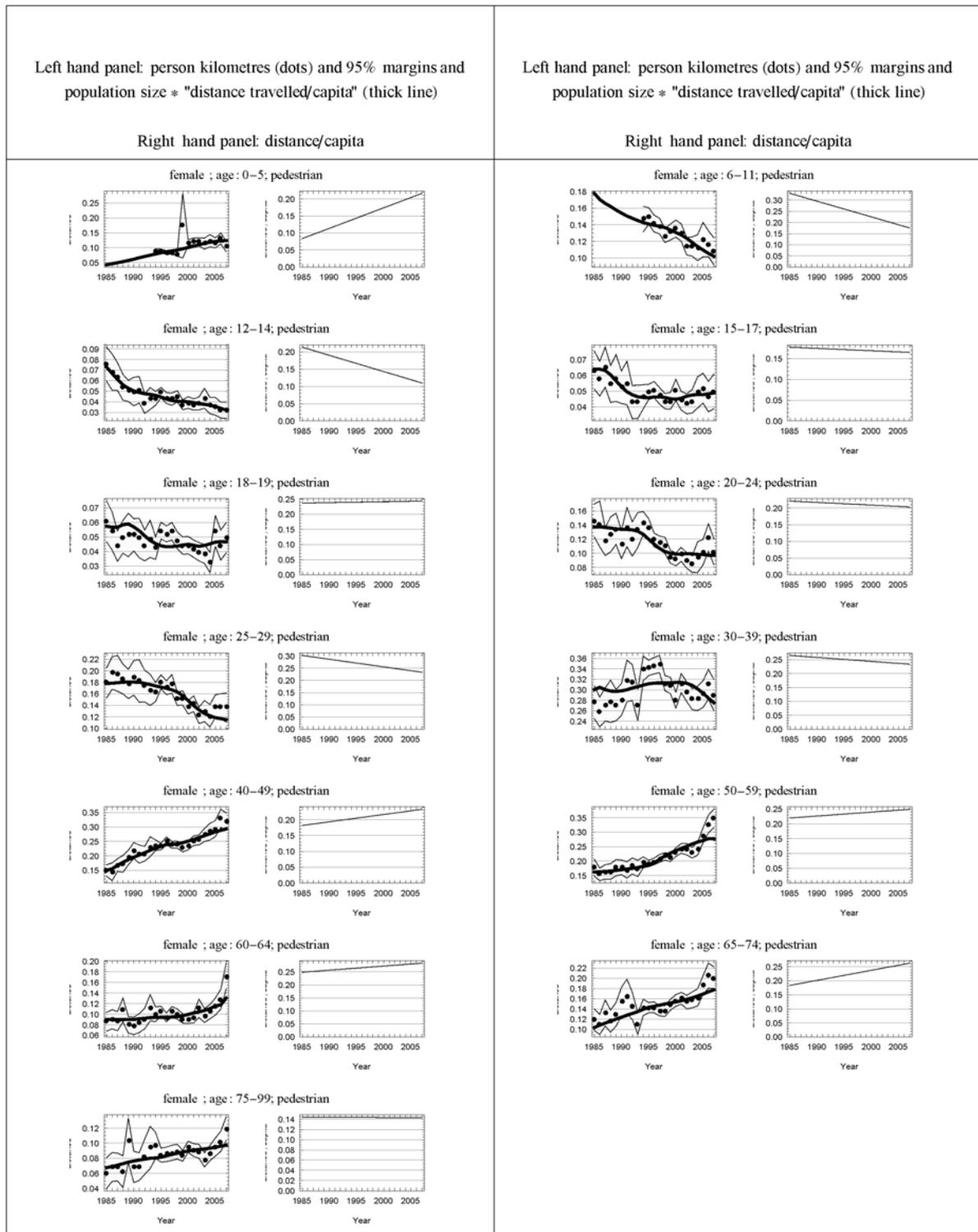


Figure A4.1.32 Annual pedestrian distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

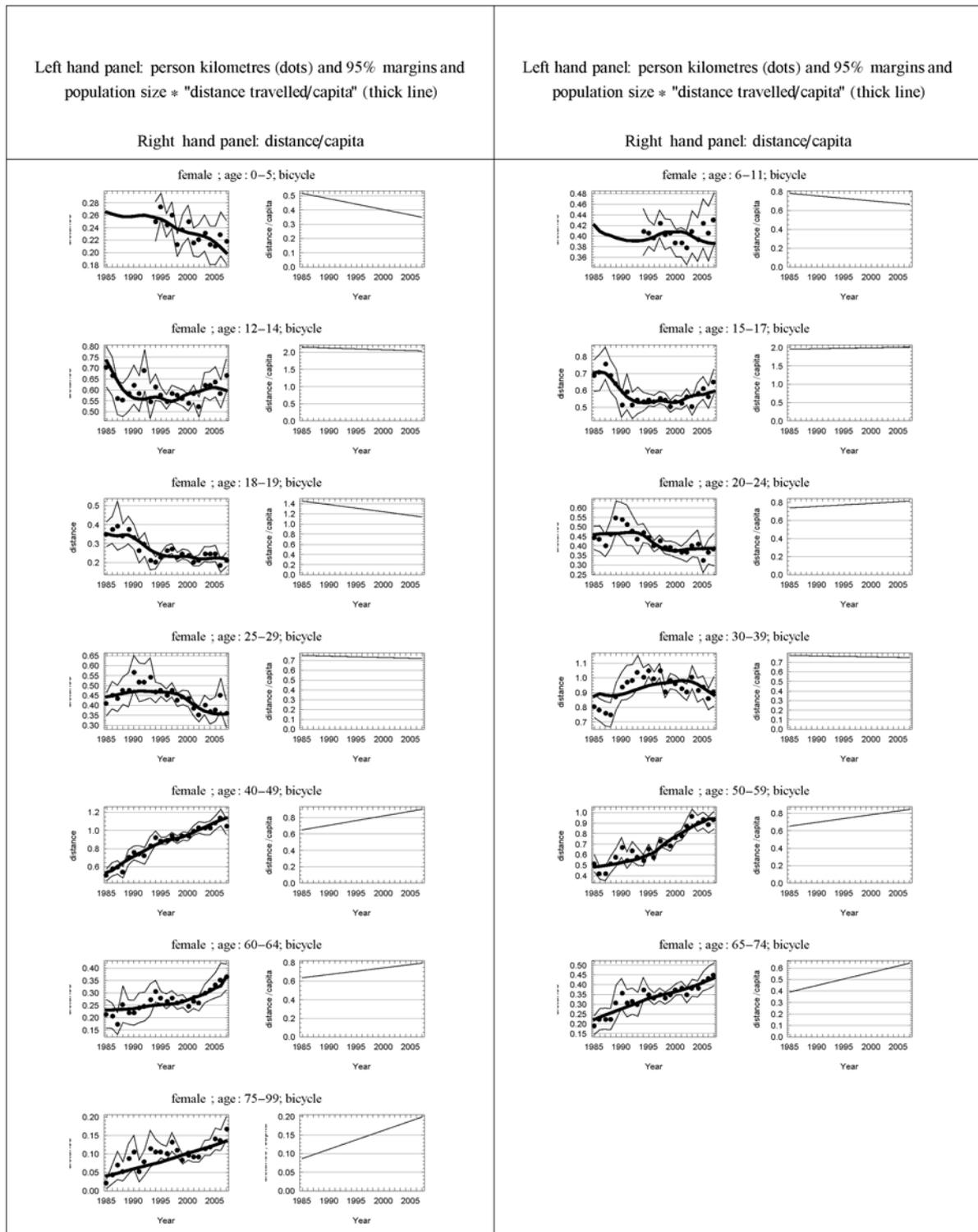


Figure A4.1.33 Annual bicycle distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

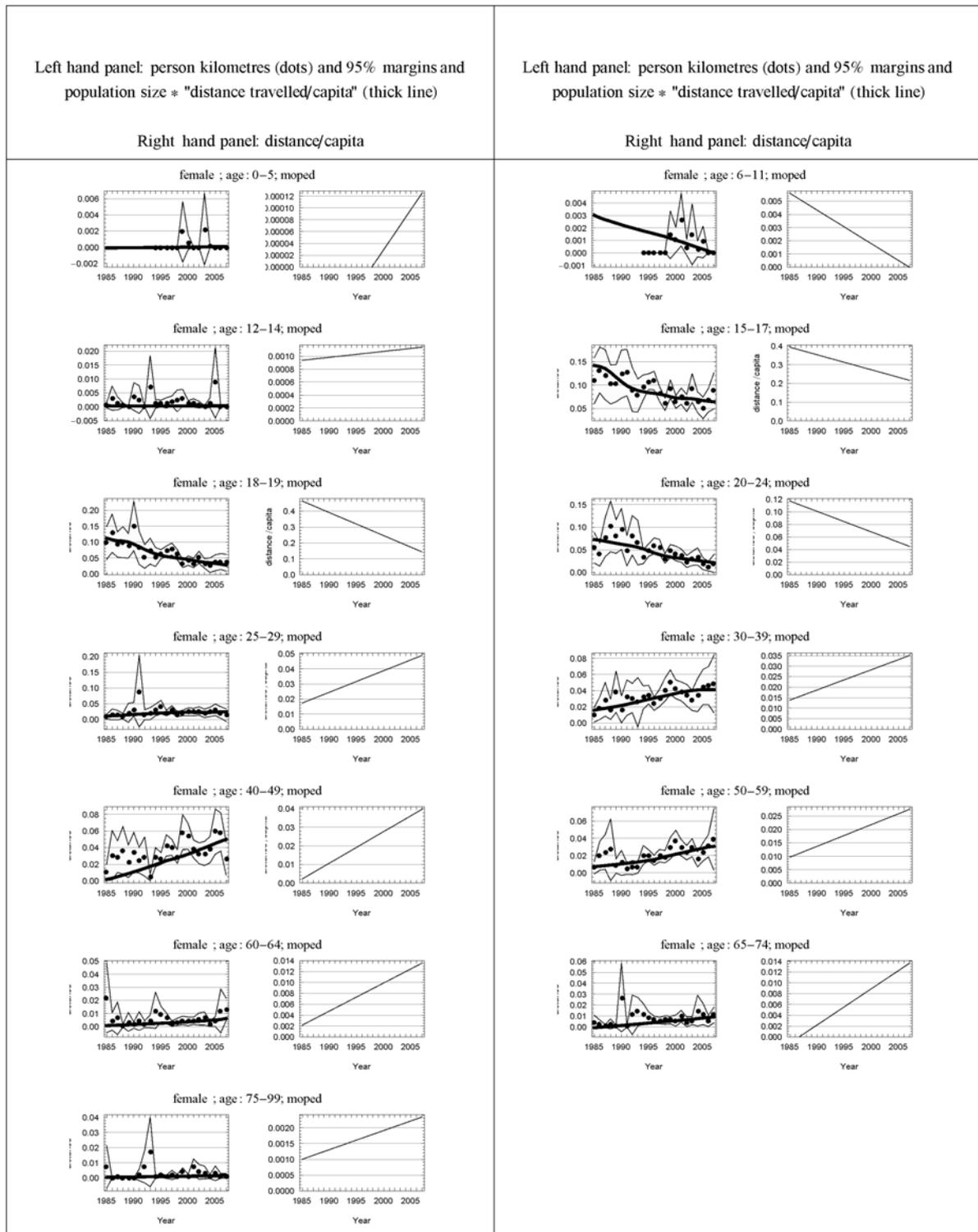


Figure A4.1.34 Annual moped distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

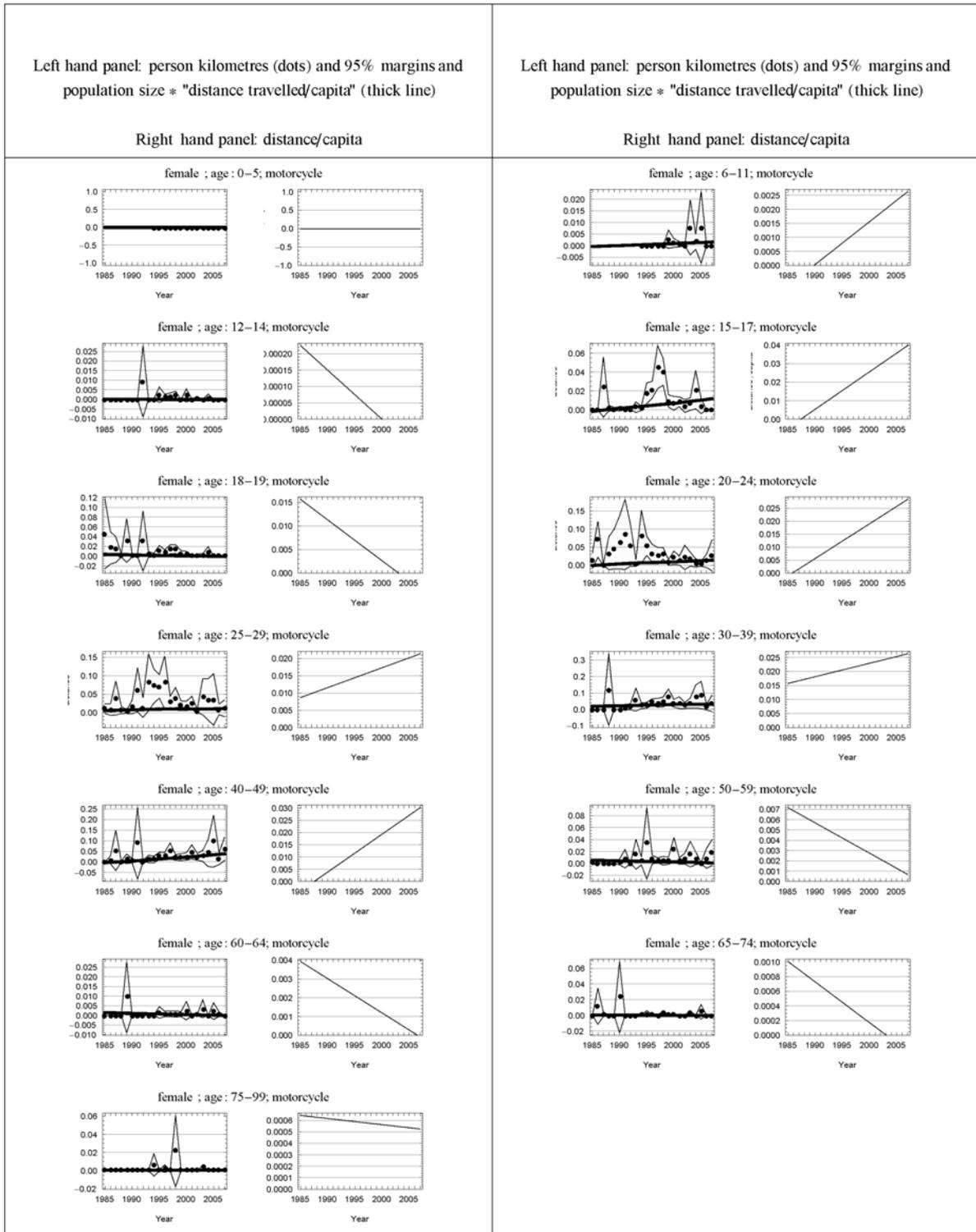


Figure A4.1.35 Annual motorcycle distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

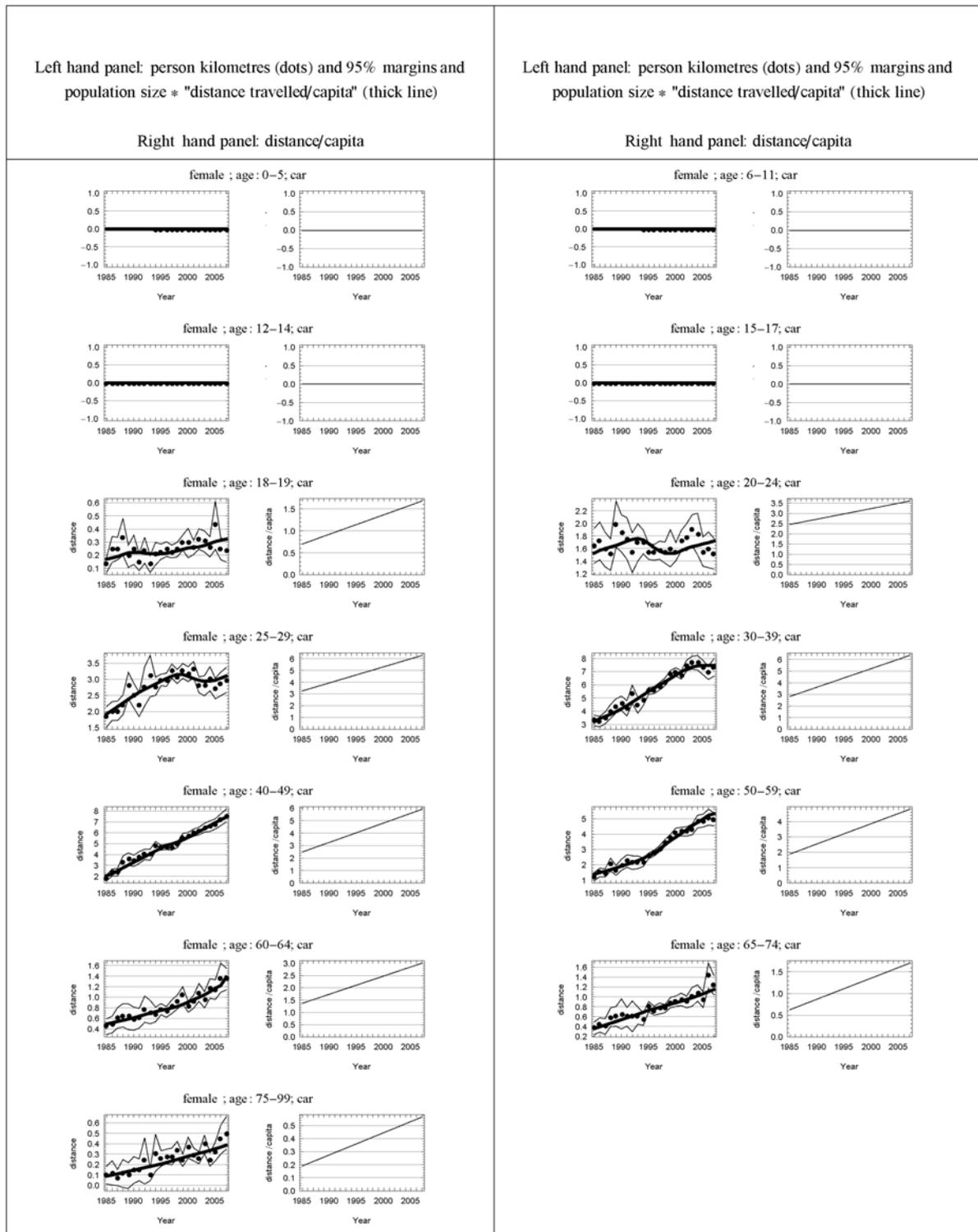


Figure A4.1.36 Annual passenger car driver distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

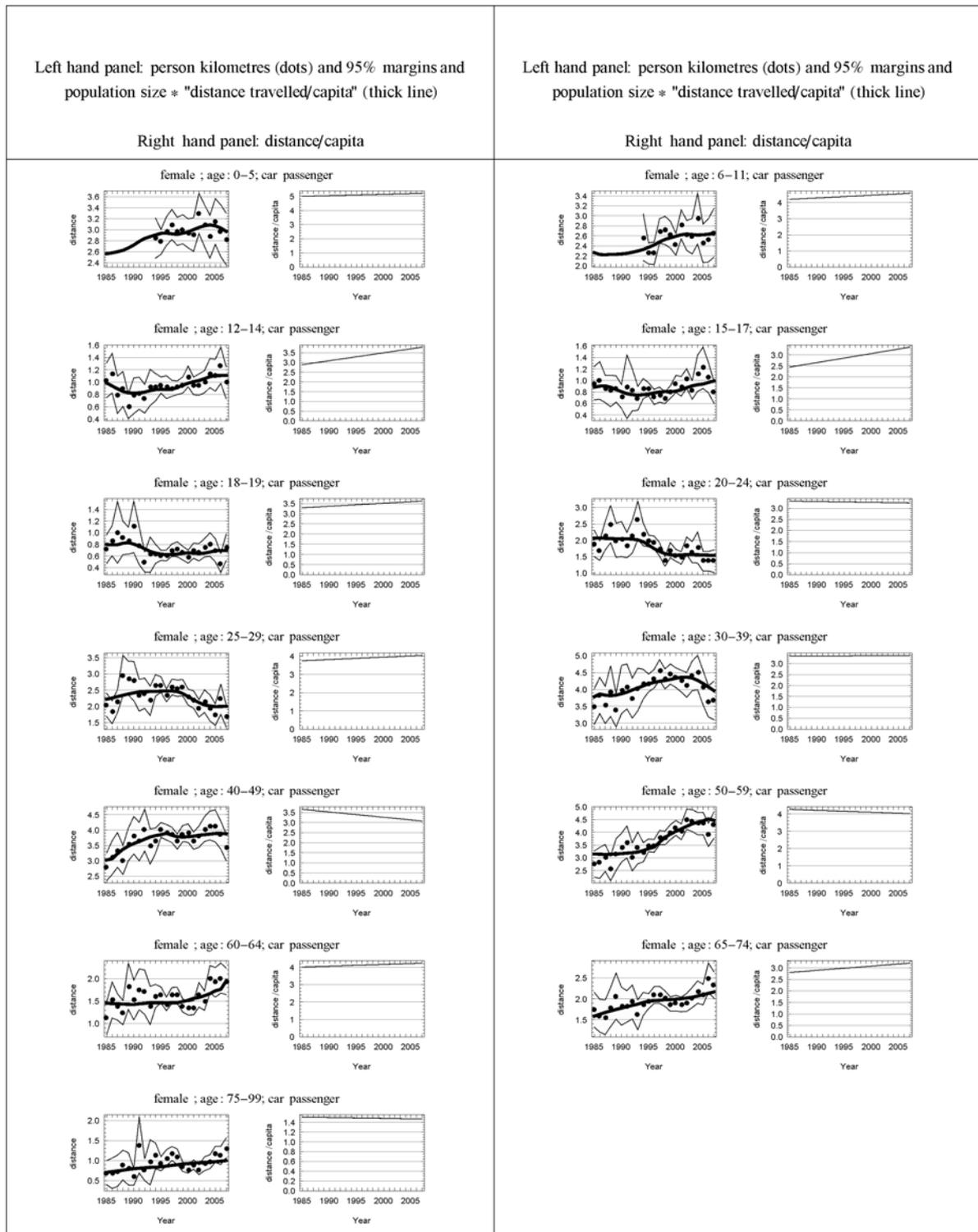


Figure A4.1.37 Annual car passenger distance travelled, by female inhabitants, by age. For further explanation see Figure A4.1.1.

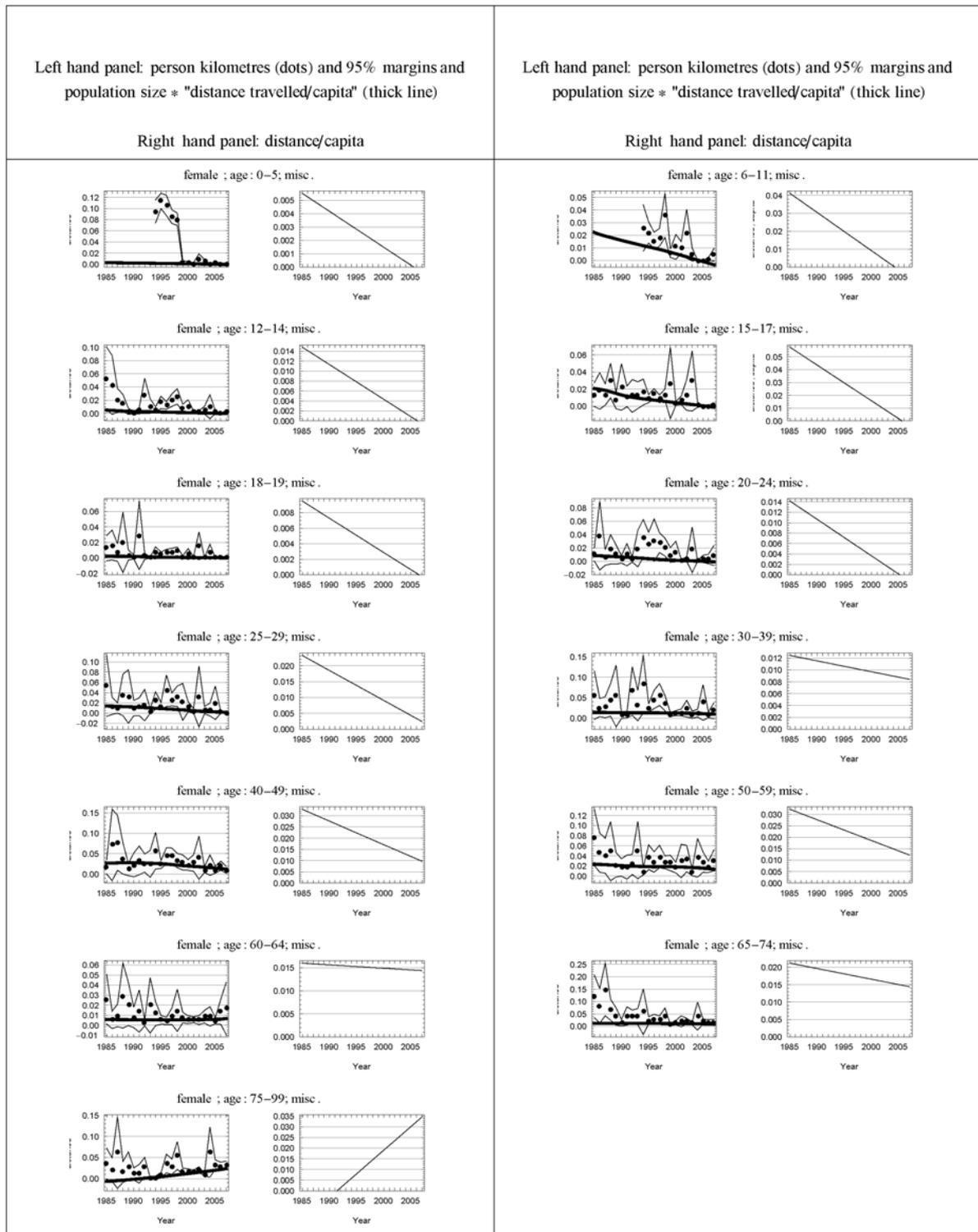


Figure A4.1.37 Annual distance travelled in miscellaneous vehicles, by female inhabitants, by age. For further explanation see Figure A4.1.1.

Appendix A4.2

Penalty function

The 'penalty' in (4.2) is a function consisting of three terms. Let us define $\tau_{k,t} = a_{ij} + b_{ij}t$ and δ_k to simplify the notation, where $k = 1, \dots, 100$ are the individual age years, and $t = 1, \dots, n$ are time years as before. Then, the penalty in (4.2) can be written as

$$L_1 p_1 + L_2 p_2 + L_3 p_3,$$

where

$$p_1 = \sum_{t=1}^n \sum_{k=1}^{100} [(\tau_{k+1,t} - \tau_{k,t}) - (\tau_{k,t} - \tau_{k-1,t})]^2,$$

$$p_2 = \sum_{k=1}^{100} [(\delta_{k+1} - \delta_k) - (\delta_k - \delta_{k-1})]^2,$$

$$p_3 = \sum_{t=1}^n \sum_{k=1}^{100} (\tau_{k,t} - \tau_{k,t-1} - \delta_k)^2,$$

are the penalty *terms*, and L_1 , L_2 , and L_3 are the penalty *parameters*. In practice, p_3 is held strictly by applying a very large value for L_3 , which resulted in a linear trend for each age year.

Appendix A4.3

Technical details of the disaggregated age-gender model

In this section a short overview of the models used in the analyses are presented. We also present the different situations that were analysed, for example different choices for (number of) age classes and covariance structure of observation and state disturbances. Basically, three different models were fitted: a model using mobility data only, a model using mobility and population size data and finally a model using only population size data.

Mathematical description of the models

In the first model, no data on the size of the population were used. Letting F_t denote the annual number of accident outcomes (which can be either fatalities or people KSI) for an age class and gender in year t , and M_t the annual number of kilometres travelled by that age class and gender of the road user population then the first model can be written in simplified form as:

$$\begin{aligned} \log(F_{i,t}) &= \mu_{i,t}^{(1)} + \mu_{i,t}^{(2)} + \varepsilon_{i,t}^{(1)} \\ \log(M_{i,t}) &= \mu_{i,t}^{(1)} + \varepsilon_{i,t}^{(2)} \end{aligned} \quad (A4.3.1a)$$

where the latent exposure $\mu_{i,t}^{(1)}$ and the latent risk $\mu_{i,t}^{(2)}$ are treated as local linear trend models:

$$\begin{aligned} \mu_{i,t+1}^{(1)} &= \mu_{i,t}^{(1)} + v_{i,t}^{(1)} + \xi_{i,t}^{(1)} \\ v_{i,t+1}^{(1)} &= v_{i,t}^{(1)} + \zeta_{i,t}^{(1)} \\ \mu_{i,t+1}^{(2)} &= \mu_{i,t}^{(2)} + v_{i,t}^{(2)} + \xi_{i,t}^{(2)} \\ v_{i,t+1}^{(2)} &= v_{i,t}^{(2)} + \zeta_{i,t}^{(2)} \end{aligned} \quad (A4.3.1b)$$

for $t = 1, \dots, n$, and the $\mu_{i,t}$ are level components and the $v_{i,t}$ are slope components. The terms $\varepsilon_{i,t}$ are observation residuals or disturbances, and the terms $\xi_{i,t}$ and $\zeta_{i,t}$ are state disturbances. The index i in (A4.3.1a) and (A4.3.1b) always indicates a specific combination of an age class and gender (e.g., male road users of 20-24 years old). Thus, if we have C age classes then there are $2C$ subgroups, and the index i in (A4.3.1a) and (A4.3.1b) runs from 1, ..., $2C$, meaning that there are actually $4C$ equations in (A4.3.1a) and $8C$ equations in (A4.3.1b).

In the second model, annual data are used both of the number of kilometres travelled *and* the population size:

$$\begin{aligned} \log(F_{i,t}) &= \mu_{i,t}^{(1)} + \log(x_{i,t}) + \mu_{i,t}^{(2)} + \varepsilon_{i,t}^{(1)} \\ \log(M_{i,t}) &= \mu_{i,t}^{(1)} + \log(x_{i,t}) + \varepsilon_{i,t}^{(2)} \end{aligned} \quad (A4.3.2)$$

where $\log(x_{i,t})$ is the logarithm of the population size in an age class/gender combination -treated as a fixed variable-. The latent exposure $\mu_{i,t}^{(1)}$ and the latent risk $\mu_{i,t}^{(2)}$ are again treated as local linear trend models.

In the third and last model only annual data of the population size are used:

$$\begin{aligned} \log(F_{i,t}) &= \log(x_{i,t}) + \mu_{i,t}^{(2)} + \varepsilon_{i,t}^{(1)} \\ \log(M_{i,t}) &= \log(x_{i,t}) + \varepsilon_{i,t}^{(2)} \end{aligned}, \quad (\text{A.4.3.3})$$

where $\log(x_{i,t})$ is again handled as a fixed variable, and the latent risk $\mu_{i,t}^{(2)}$ is treated as a local linear trend model.

In all three models, which are referred to as model A4.3.1, model A4.3.2, and model A4.3.3, respectively, six different (co)variance structures were applied to the observation and state disturbances. The reason why we applied all these covariance structures is that the multivariate analysis of time series with unobserved component models by state space methods is a relatively new, and as yet mostly unexplored field. The first covariance structure represents the most independent cases (a completely deterministic analysis), the sixth covariance structures the case in which most interaction between the development is allowed.

In the first covariance structure, the variances of all observation disturbances were estimated but they were all assumed to be uncorrelated. Moreover, all state disturbances variances and co-variances were assumed to be zero, which is identical to a multivariate classical linear regression model. In the second covariance structure all variances of the observation and state disturbances were estimated (resulting in local linear trend models for the state components) but they were all assumed to be uncorrelated (not affecting each other). The third covariance structure was identical to the second one, except that the latent exposures and risks were now allowed to be correlated *within each age class but not within genders of that age class*. The fourth covariance structure was identical to the third one, except that the latent exposures and risks were not only allowed to be correlated *within each age class but also within both genders of that age class*.

The fifth covariance structure was identical to the second one, except that all latent risks were allowed to be correlated *both within age classes and genders*, and finally in the sixth and last covariance structure *all* state components were allowed to be correlated, irrespective of risk, exposure, age class or gender. Moreover, in the last and sixth covariance structure all observation or measurement disturbances were also allowed to covariate. The covariance structures considered are further detailed in the next section.

On top of all these combinations we also explored the use of

1. fatalities per age class and gender combination versus

2. persons KSI per age class and gender combination, *and* considered the following numbers of age classes: two age classes (both 0-49 year old road users versus 50+ year old road users, and 0-24 old road users versus 25+ old road users), three age classes (road users of 0-19 years old, 20-64 years old, and 65+ years old), four age classes (road users of 0-14 years old, 15-24 years old, 25-64 years old, and 65+ years old), and five age classes (road users of 0-14 years old, 15-19 years old, 20-29 years old, 30-64 years old, and 65+ years old).

Finally, all these analyses were performed using the following three approaches: forecasts until 2040 from the year 2004 onwards, forecasts until 2040 from the year 2007 onwards excluding a level shift intervention variable for the risks in 2004, and forecasts until 2040 from the year 2007 onwards *including* a level shift intervention variable for the risks in 2004. The motivation for investigating the latter two options is that a significant drop in the total number of fatalities has been observed in the Netherlands in 2004, see Stipdonk et al. (2006).

Covariance structure matrices

We will illustrate the (co)variance structures for the most simple case of two age classes only where we let a1 and a2 denote these two age classes, and g1 and g2 the two genders (males and females) of road users. Furthermore we use the notation 'exp' for the latent exposure $\mu_{i,t}^{(1)}$ of a subgroup, 'risk' its latent risk $\mu_{i,t}^{(2)}$, and 'var' the variance of the corresponding state disturbances.

The first covariance matrix is the zero matrix.

The second (co)variance matrix of the state disturbances has the following structure:

	a1g1 (exp)	a1g1 (risk)	a1g2 (exp)	a1g2 (risk)	a2g1 (exp)	a2g1 (risk)	a2g2 (exp)	a2g2 (risk)
a1g1 (exp)	var	0	0	0	0	0	0	0
a1g1 (risk)	0	var	0	0	0	0	0	0
a1g2 (exp)	0	0	var	0	0	0	0	0
a1g2 (risk)	0	0	0	var	0	0	0	0
a2g1 (exp)	0	0	0	0	var	0	0	0
a2g1 (risk)	0	0	0	0	0	Var	0	0
a2g2 (exp)	0	0	0	0	0	0	var	0
a2g2 (risk)	0	0	0	0	0	0	0	var

The third covariance structure can be written as:

	a1g1 (exp)	a1g1 (risk)	a1g2 (exp)	a1g2 (risk)	a2g1 (exp)	a2g1 (risk)	a2g2 (exp)	a2g2 (risk)
a1g1 (exp)	var	covar	0	0	0	0	0	0
a1g1 (risk)	covar	var	0	0	0	0	0	0
a1g2 (exp)	0	0	var	covar	0	0	0	0
a1g2 (risk)	0	0	covar	var	0	0	0	0
a2g1 (exp)	0	0	0	0	var	covar	0	0
a2g1 (risk)	0	0	0	0	covar	var	0	0
a2g2 (exp)	0	0	0	0	0	0	var	covar
a2g2 (risk)	0	0	0	0	0	0	covar	var

The fourth covariance structure can be written as:

	a1g1 (exp)	a1g1 (risk)	a1g2 (exp)	a1g2 (risk)	a2g1 (exp)	a2g1 (risk)	a2g2 (exp)	a2g2 (risk)
a1g1 (exp)	var	covar	covar	covar	0	0	0	0
a1g1 (risk)	covar	var	covar	covar	0	0	0	0
a1g2 (exp)	covar	covar	var	covar	0	0	0	0
a1g2 (risk)	covar	covar	covar	var	0	0	0	0
a2g1 (exp)	0	0	0	0	var	covar	covar	covar
a2g1 (risk)	0	0	0	0	covar	var	covar	covar
a2g2 (exp)	0	0	0	0	covar	covar	var	covar
a2g2 (risk)	0	0	0	0	covar	covar	covar	var

The fifth covariance structure can be written as:

	a1g1 (exp)	a1g1 (risk)	a1g2 (exp)	a1g2 (risk)	a2g1 (exp)	a2g1 (risk)	a2g2 (exp)	a2g2 (risk)
a1g1 (exp)	var	0	0	0	0	0	0	0
a1g1 (risk)	0	var	0	covar	0	covar	0	covar
a1g2 (exp)	0	0	var	0	0	0	0	0
a1g2 (risk)	0	covar	0	var	0	covar	0	covar
a2g1 (exp)	0	0	0	0	var	0	0	0
a2g1 (risk)	0	covar	0	covar	0	var	0	covar
a2g2 (exp)	0	0	0	0	0	0	var	0
a2g2 (risk)	0	covar	0	covar	0	covar	0	var

The sixth covariance structure can be written as:

	a1g1 (exp)	a1g1 (risk)	a1g2 (exp)	a1g2 (risk)	a2g1 (exp)	a2g1 (risk)	a2g2 (exp)	a2g2 (risk)
a1g1 (exp)	var	covar	covar	covar	covar	covar	covar	covar
a1g1 (risk)	covar	var	covar	covar	covar	covar	covar	covar
a1g2 (exp)	covar	covar	var	covar	covar	covar	covar	covar
a1g2 (risk)	covar	covar	covar	var	covar	covar	covar	covar
a2g1 (exp)	covar	covar	covar	covar	var	covar	covar	covar
a2g1 (risk)	covar	covar	covar	covar	covar	var	covar	covar
a2g2 (exp)	covar	covar	covar	covar	covar	covar	var	covar
a2g2 (risk)	covar	covar	covar	covar	covar	covar	covar	var

Appendix A4.4

We will first compare the possible effects of using different modelling approaches on the forecasts. Later the use of population figures is introduced. All graphs contain forecasts up to 2040. This is a very long time ahead from 2008, and arguably too long to assume a useful forecast. The reason to include a very long term prognosis is that it helps reveal differences in prognosis from the three models.

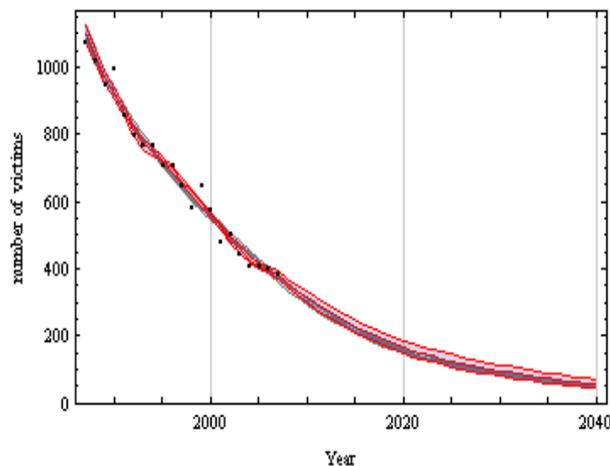


Figure A4.4.1. Forecasts from 2007 onwards) of the number of victims of male KSI in age class 0-14 as obtained with model 2 using 4 age classes including 68% confidence intervals, first versus sixth covariance structure.

In Figure A4.4.1, the forecast for male KSI in age class 0-14 of model 2 obtained with the first and fixed covariance structure (deterministic analysis) is compared with the forecast for the same data and model using the full and sixth covariance structure (a completely stochastic analysis). The resulting forecasts and their confidence intervals are almost identical, even though these two modelling approaches are completely different. This is because these data display a rather smooth exponential development, which is recovered even with a completely deterministic model.

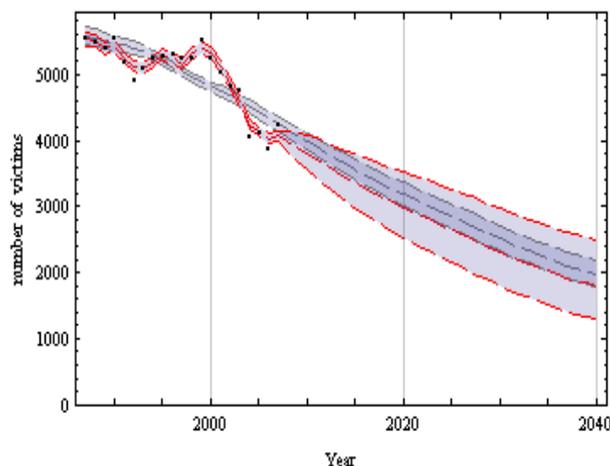


Figure A4.4.2. Forecasts from 2007 onwards of male KSI in age class 20-64 as obtained with model 3 using 3 age classes including 68% confidence intervals, first (dark region) versus sixth (light region) covariance structure.

In Figure A4.4.2, the forecasts for male KSI in age class 20-64 of model 3 obtained with the first and fixed covariance structure (deterministic forecast) are compared with the forecasts for the same data and model using the full and sixth covariance structure (a completely stochastic analysis). In contrast with the previous figure we obtain different forecasts. Also, the uncertainty for the sixth covariance structure is larger than that for the first covariance structure. This reflects that deterministic trends are not appropriate for modelling this particular development, yielding a false sense of certainty concerning the forecasts. Stochastic trends on the other hand are appropriate in this case, and result in proper confidence intervals for the forecasts.

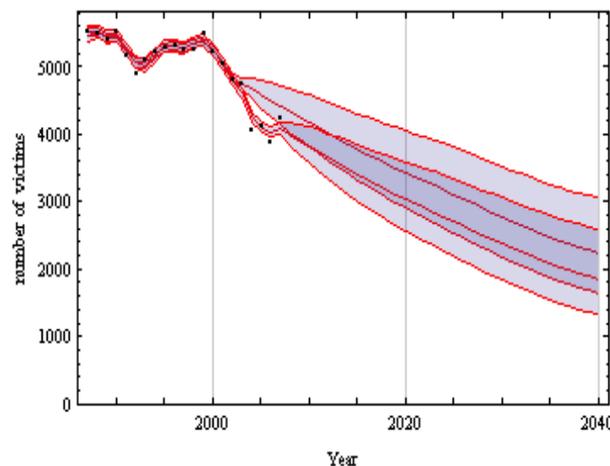


Figure A4.4.3. Forecasts from 2003 onwards (first, third and fifth line from above) and from 2007 onwards (second, fourth and sixth line from above) without intervention variables for risk in 2004 of males KSI in age class 20-64 as obtained with model 2 using 3 age classes and sixth covariance structure, including 68% confidence intervals.

Figure A4.4.3 illustrates for males KSI of 20-64 years old what difference it can make whether the forecasts are based on the time period 1985-2003 or on the time period 1985-2007 (that is including the sudden drop in fatalities in 2004).

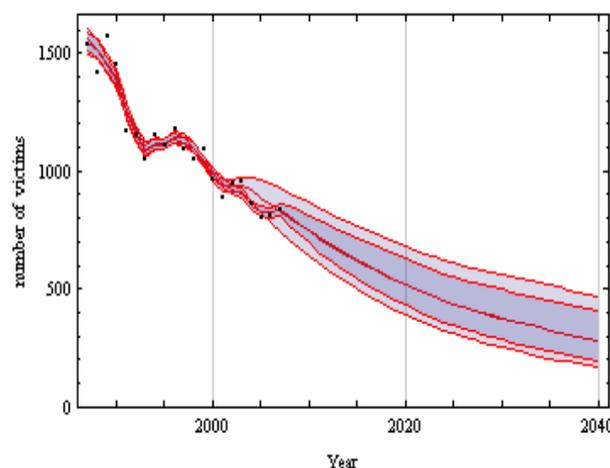


Figure A4.4.4. Forecasts from 2003 onwards (light region) versus from 2007 (dark region) onwards including intervention variables for risk in 2004 of female KSI in age class 0-19 as obtained with model 2 using 3 age classes and sixth covariance structure, including 68% confidence intervals.

In Figure A4.4., on the other hand, where forecasts for females KSI of 0-19 years old are compared in terms of whether they are based on the time period 1985-2003 or on the time period 1985-2007, hardly any differences are found. Clearly the drop in casualties in 2004 does not apply to this subgroup of road users!

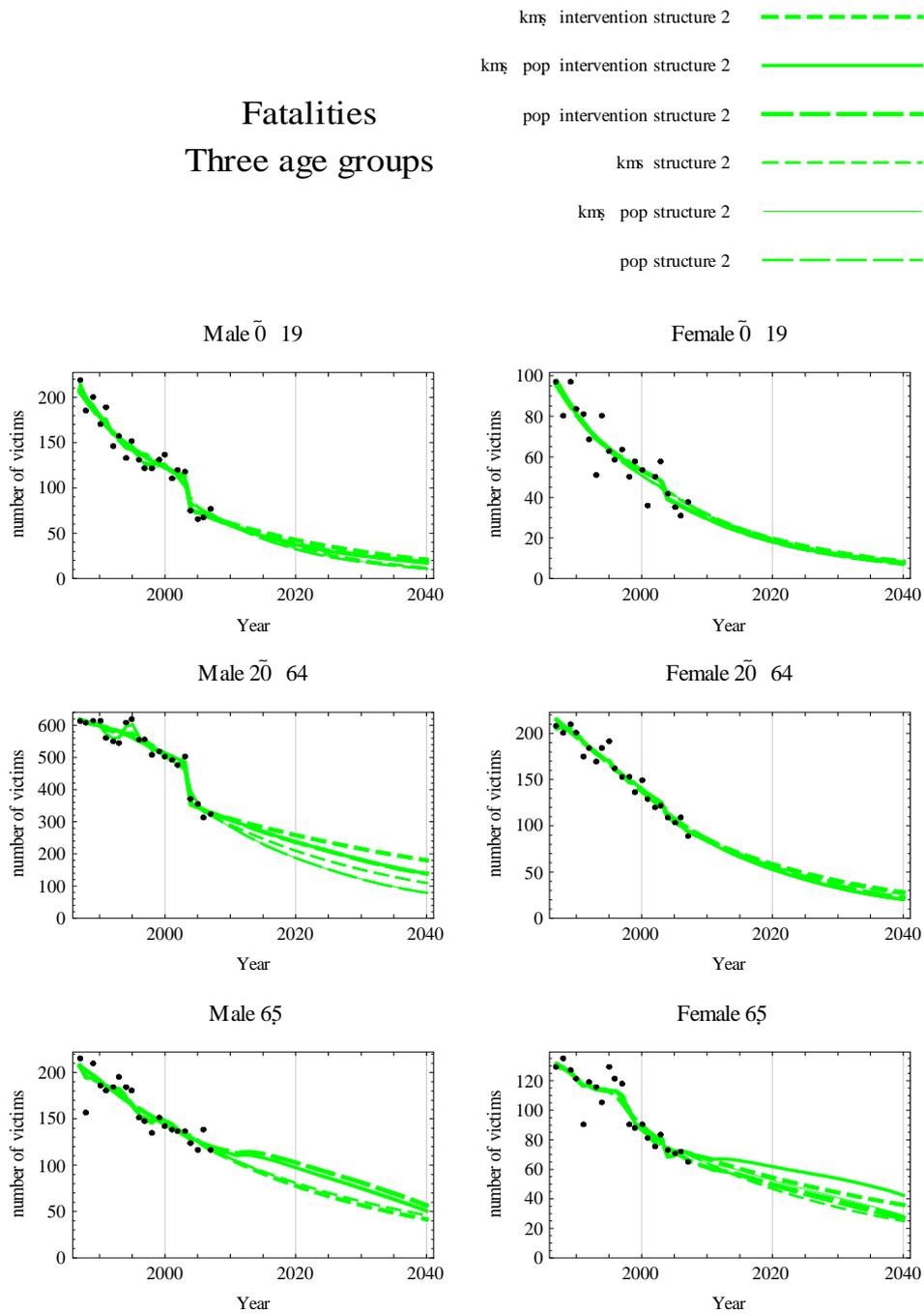


Figure A4.4.5. Forecasts from 2007 onwards of fatalities in 3 age classes as obtained with model 1, model 2, and model 3 with third covariance structure, both including and excluding intervention variables for risks in 2004.

In Figure A4.4.5 and A4.4.6 we focus on the effects of including data on population size in the estimation of disaggregated forecasts until 2040, and compare the results obtained with model 1 (mobility figures only), model 2 (both mobility and population figures), and model 3 (population figures only). The figures constitute a selection of many results that were found.

In Figure A4.4.5, the forecasts obtained with these three models are compared for the fatalities disaggregated in three age classes based on data for the years 1985-2007 including and excluding intervention variables for the risks in 2004, and using the same third covariance structure mentioned above in all cases. The thick lines are the forecasts obtained with the three models including intervention variables for 2004, the thin lines are the forecasts obtained without these intervention variables.

First of all, we see that the inclusion of population figures in the analysis is especially relevant for the forecasts of males in the 20-64 age groups, and for the males and females in the 65+ age groups, since it is for these subgroups that the largest differences in forecasts between the three models are found.

We also see that there were significant drops in the fatalities in 2004 for the males of 0-19 years old and 20-64 years old, but not for the other four subgroups. The inclusion of a level shift intervention variable for the risk in 2004 is especially relevant for the male fatalities in the 20-64 years group, since the forecasts in- and excluding intervention variables in 2004 are markedly different for this subgroup.

In Figure A4.4.6 the same comparisons are made as in Figure A4.4.5, but now for the situation where four age classes are distinguished instead of three; here we only show the analyses including intervention variables applied to the risks in 2004.

In this case the inclusion of population figures in the analyses is found to be especially relevant for the forecasts of male fatalities in the 15-24, 25-64 and 65+ categories, and for the forecasts of the female fatalities in the 25-64 and 65+ categories. The latter case is an interesting example, as model 2 (with both mobility data and population data) behaves differently from model 1 and model 3. To a different and lesser extent this is also the case for females 15-24. Not including the intervention in 2004 reduces the differences. The same analysis performed using killed or seriously injured victims, with or without an intervention in 2004, reveals the general pattern of model 2 and model 3 resulting in similar prognosis and model 1 resulting in a different prognosis.

Finally, only the 0-14, 15-24 and 25-64 year old males and the 15-24 year old females display significant reductions in fatalities in 2004.

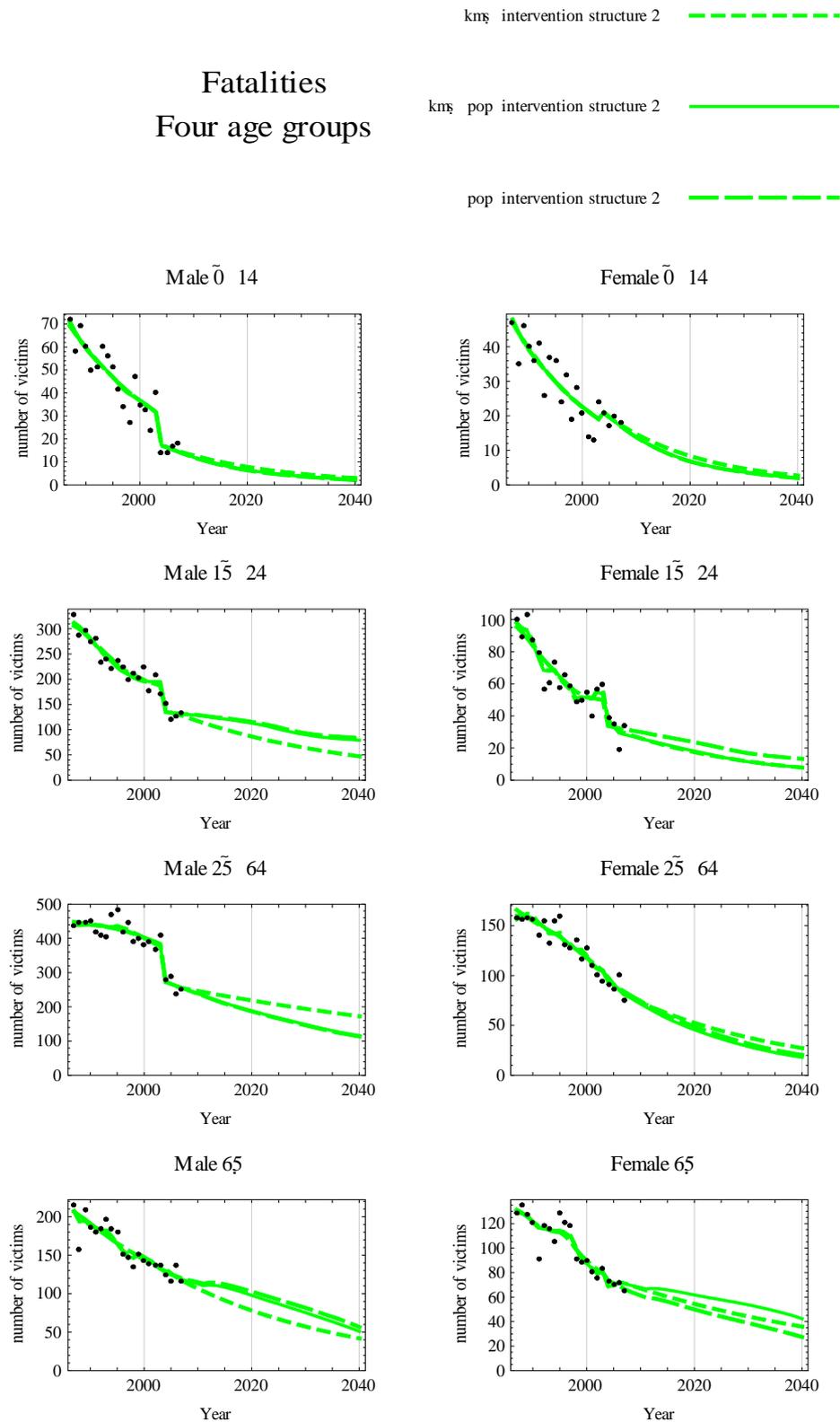


Figure A4.4.6. Forecasts from 2007 onwards of fatalities in 4 age classes as obtained with model 1, model 2, and model 3 with third covariance structure, including intervention variables for risks in 2004.

Appendix A5.1

	Estimate	Std. Error	z value
(Intercept)	2.975	0.024	122.274
5<MeanTemp<10 °C	0.194	0.026	7.550
10<MeanTemp<20 °C	0.280	0.025	11.389
20<MeanTemp<30 °C	0.334	0.025	13.353
MeanTemp>30 °C	0.385	0.030	12.959
0<TotalPrecipitation<5mm	-0.024	0.010	-2.320
5<TotalPrecipitation<10mm	-0.073	0.023	-3.180
10<TotalPrecipitation<20	-0.097	0.023	-4.226
20<TotalPrecipitation<40	-0.190	0.031	-6.038
40<TotalPrecipitation<60	-0.458	0.074	-6.147
TotalPrecipitation>60mm	-0.557	0.103	-5.413
June	0.068	0.013	5.288
August	-0.257	0.013	-19.126
OctNovDec	0.047	0.007	6.511
Null deviance:	8754.5	(7659 d.o.f.)	
Residual deviance:	7768.4	(7646 d.o.f.)	
AIC:	51847		

Table A5.1: Model specification and estimation results for accidents (daily data)

	Estimate	Std. Error	z value
(Intercept)	-0.23	0.10	-2.33
5<MeanTemp<10 °C	0.28	0.10	2.68
10<MeanTemp<20 °C	0.32	0.10	3.23
20<MeanTemp<30 °C	0.38	0.10	3.86
MeanTemp>30 °C	0.36	0.12	3.08
TotalPrecipitation>60mm	-1.01	0.52	-1.93
Null deviance:	8522.3	(7659 d.o.f.)	
Residual deviance:	8494.8	(7654 d.o.f.)	
AIC:	21691		

Table A5.2: Model specification and estimation results for fatalities (daily data)

	Estimate	Std.	Error
(Intercept)	6.829	0.096	71.167
MinMeanTemp<5 °C	-0.240	0.046	-5.196
SumTotalPrecipitation	-0.001	0.000	-2.213
Passenger Cars	-0.001	0.000	-6.766
Heavy Trucks	0.010	0.003	3.369
Null deviance:	149.19	(76 d.o.f.)	
Residual deviance:	77.33	(72 d.o.f.)	
AIC:	950.53		

Table A5.3: Model specification and estimation results for accidents (monthly data – GLM)

Appendix A5.2

The Structural time series model with explanatory variables (or unobserved components model)

In this appendix, we are interested by the multivariate structural times series model - or *unobserved components model* - with explanatory variables, which is used in section 5.3, for modelling the development of injury accidents data, aggregated at a region's level, whether France, the Netherlands or the Athens region, and taking into account explanatory variables measuring meteorological factors.

The structural time series model is a particular case of state space model (of which one application was given in the preceding section 3.2). As such, it is often presented under a *state space form*, which is defined by the two following observation and state equations:

$$y_t = \left(\begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \right) x_t + D(B)z_t + \varepsilon_t$$

and

$$x_t = \left(\begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \right) x_{t-1} + \kappa_t$$

where,

$$x_t = \left[\mu_{1,t}, \nu_{1,t}, \gamma_{1,t}, \dots, \gamma_{1,t-s+1}, \dots, \mu_{m,t}, \nu_{m,t}, \gamma_{m,t}, \dots, \gamma_{m,t-s+1} \right]'$$

$$y_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{m,t} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{m,t} \end{bmatrix}, z_t = \begin{bmatrix} z_{1t} \\ \vdots \\ z_{kt} \end{bmatrix}$$

$$\kappa_t = \left[\eta_{1,t}, \zeta_{1,t}, \omega_{1,t}, 0, \dots, 0, \dots, \eta_{m,t}, \zeta_{m,t}, \omega_{m,t}, 0, \dots, 0 \right]'$$

here \otimes denotes the Kronecker product and κ_t the perturbation of the state equation.

The k explanatory variables are taken into account through the vector z_t , in the observation equation.

In practice, the *scalar form* of the model is often preferred, as it highlights the structure of each of the *unobserved components*: the level and the slope of the trend, and the seasonal component: μ , β and γ .

$$\left\{ \begin{array}{l} y_{1,t} = \mu_{1,t} + \gamma_{1,t} + \sum_i^k d_{1,i} z_{i,t} + \varepsilon_{1,t} \\ y_{2,t} = \mu_{2,t} + \gamma_{2,t} + \sum_i^k d_{2,i} z_{i,t} + \varepsilon_{2,t} \\ \vdots \\ y_{N,t} = \mu_{N,t} + \gamma_{N,t} + \sum_i^k d_{N,i} z_{i,t} + \varepsilon_{N,t} \end{array} \right. \quad \left\{ \begin{array}{l} \mu_{1,t+1} = \mu_{1,t} + \beta_{1,t} + \xi_{1,t} \\ \mu_{2,t+1} = \mu_{2,t} + \beta_{2,t} + \xi_{2,t} \\ \vdots \\ \mu_{N,t+1} = \mu_{N,t} + \beta_{N,t} + \xi_{N,t} \end{array} \right. \quad N,t$$

$$\left\{ \begin{array}{l} \beta_{1,t+1} = \beta_{1,t} + \varsigma_{1,t} \\ \beta_{2,t+1} = \beta_{2,t} + \varsigma_{2,t} \\ \vdots \\ \beta_{N,t+1} = \beta_{N,t} + \varsigma_{N,t} \end{array} \right. \quad \left\{ \begin{array}{l} \gamma_{1,t} = -\sum_{i=1}^{s-1} \gamma_{1,t-i} + \omega_{1,t} \\ \gamma_{2,t} = -\sum_{i=1}^{s-1} \gamma_{2,t-i} + \omega_{2,t} \\ \vdots \\ \gamma_{N,t} = -\sum_{i=1}^{s-1} \gamma_{N,t-i} + \omega_{N,t} \end{array} \right.$$

The variance-covariance matrices of the disturbances ε_t and κ_t are given by

$$\text{Var}_{\varepsilon_t} = \begin{bmatrix} \text{Var}_{\varepsilon_{1,t}} & \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) & \dots & \text{Cov}(\varepsilon_{1,t}, \varepsilon_{m,t}) \\ \text{Cov}(\varepsilon_{2,t}, \varepsilon_{1,t}) & \text{Var}_{\varepsilon_{2,t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \text{Cov}(\varepsilon_{m,t}, \varepsilon_{1,t}) & \dots & \text{Cov}(\varepsilon_{m,t}, \varepsilon_{m-1,t}) & \text{Var}_{\varepsilon_{m,t}} \end{bmatrix}$$

and

$$\text{Var}_{\kappa_t} = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & C & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with

$$B = \begin{bmatrix} \text{Var}_{\eta_{1,t}} & \text{Cov}(\eta_{1,t}, \eta_{2,t}) & \dots & \text{Cov}(\eta_{1,t}, \eta_{m,t}) \\ \text{Cov}(\eta_{2,t}, \eta_{1,t}) & \text{Var}_{\eta_{2,t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \text{Cov}(\eta_{m,t}, \eta_{1,t}) & \dots & \text{Cov}(\eta_{m,t}, \eta_{m-1,t}) & \text{Var}_{\eta_{m,t}} \end{bmatrix}$$

$$C = \begin{bmatrix} \text{Var}^{\xi_{1,t}} & \text{Cov}(\xi_{1,t}, \xi_{2,t}) & \dots & \text{Cov}(\xi_{1,t}, \xi_{m,t}) \\ \text{Cov}(\xi_{2,t}, \xi_{1,t}) & \text{Var}^{\xi_{2,t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\xi_{m-1,t}, \xi_{m,t}) \\ \text{Cov}(\xi_{m,t}, \xi_{1,t}) & \dots & \text{Cov}(\xi_{m,t}, \xi_{m-1,t}) & \text{Var}^{\xi_{m,t}} \end{bmatrix}$$

and

$$D = \begin{bmatrix} \text{Var}^{\omega_{1,t}} & \text{Cov}(\omega_{1,t}, \omega_{2,t}) & \dots & \text{Cov}(\omega_{1,t}, \omega_{m,t}) \\ \text{Cov}(\omega_{2,t}, \omega_{1,t}) & \text{Var}^{\omega_{2,t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\omega_{m-1,t}, \omega_{m,t}) \\ \text{Cov}(\omega_{m,t}, \omega_{1,t}) & \dots & \text{Cov}(\omega_{m,t}, \omega_{m-1,t}) & \text{Var}^{\omega_{m,t}} \end{bmatrix}$$

In section 5.3, the scalar form is used in a univariate case when applied to a single dataset, such as the number of injury accidents for France, the Netherlands and the Athens region.

It is used in a bivariate case when applied to a couple of datasets, such as the number of injury accidents on motorways and main roads in France, and as the number of injury accidents on motorways and rural roads in the Netherlands.

At the most, six explanatory variables were used (four in case only weather variables were used, and five to six in case an exposure to risk variable was added).

Packages¹

Three main packages are used for performing a structural time series analysis:

- the R-package sspir (<http://cran.r-project.org/web/packages/sspir/index.html>),
- the procedure UCM in SAS V9 (<http://support.sas.com/rnd/app/examples/ets/melanoma/index.htm>),
- the Ox/SSfpack 2.2 (and 3) (<http://www.ssfpack.com/download.html>).

The two first packages are limited to the univariate analysis, whereas the third one performs multivariate analysis,

In the univariate framework, the UCM procedure in SAS is easy and simple to use, and systematically provides with the model's estimation the results of the tests of significance of the nature of the unobserved components (whether deterministic or stochastic), and forecasts.

The Ox/SSfpack package is the only package which allows performing a vectorial analysis: in this multivariate framework, the covariance between several variables that would not be captured with a univariate analysis of each of the variable.

Ox/SSfpack was used for performing the structural time series analyses presented in section 5.3 (univariate and bivariate cases), and the related numerical estimations.

¹ Additional information about the practical implementation of state space modelling is available at the authors (Bergel, R., Debbbarh, M.: Main packages available for state space analysis, 2008)

Data used

The following datasets were used:

- the number of injury accidents, for France (whole of France, motorways and main roads), for the Netherlands (whole of the Netherlands, motorways and rural roads), for the Athens region);
- an aggregate measure of exposure to risk, or a proxy :
 - oil sales for the whole of France (Carbu), and the number of vehicle-kilometres on motorways and main roads (PaAC, PaRN);
 - the number of driver-kilometres measured by the MON survey for the Netherlands (Driverkm);
 - the number of vehicles measured at the Shimatari toll station in the North of Athens, for the Athens region (Totalveh);
- aggregated meteorological variables, measuring :
 - the Maximum temperature of the day (0,1°C),
 - the occurrence of frost (0/1),
 - the rainfall height (mm);

The monthly meteorological variables were aggregated the following way:

For France and for the Netherlands, daily variables were first calculated by averaging daily data measured at meteorological observation points spread over the whole territory, and were then aggregated - summed or averaged - over the month, in order to construct the monthly variables.

For the Athens region, the daily variables were measured at one single observation point of the region.

In practice, four monthly variables were used (the temperature variable was split in two variables, according to the two half-years) which are noted:

- TE ; the temperature in the summer (April to September),
- TH: the temperature in the winter (October to March),
- NGEL : the number of days of frost,
- HPLUI: the rainfall height

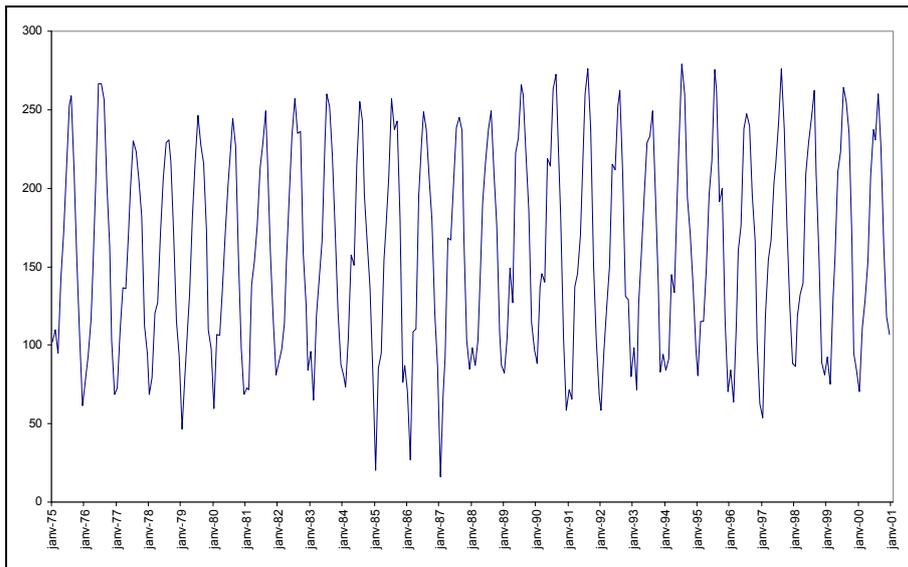


Figure A5.1: Temperature in France (0,1°C) for months from 1975 to 2000

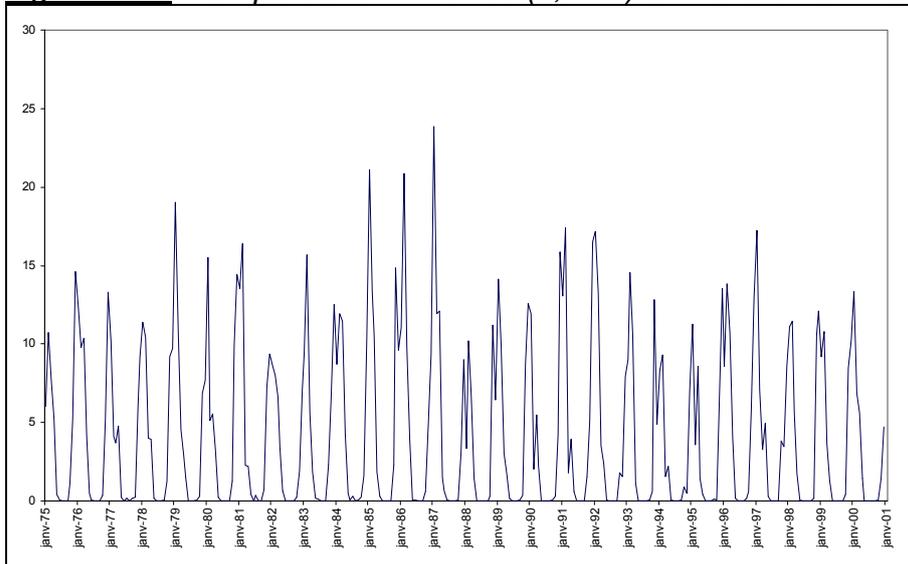


Figure A5.2: Occurrence of frost in France (days) for months from 1975 to 2000

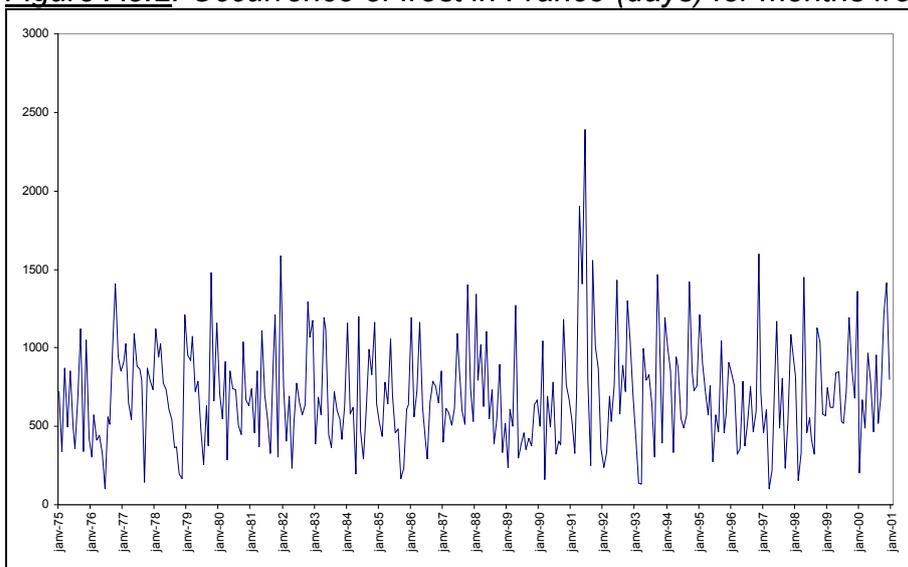


Figure A5.3: Rainfall height in France (mm) for months from 1975 to 2000

The whole of France

Numerical estimations for the model of the log number of injury accidents, for the period 1975-2000

Error variances
 level error variance is 0.000279132
 slope error variance is 2.04175e-080
 seasonal error variance is 1.41291e-006
 irregular variance is 0.000920234

Regression results

	coefficient	standard error	t-value	p-value
TE	0,0007	0,0002	3,2459	0,0013
TH	0,0012	0,0003	4,1969	0,0000
HPLUI	0,0000	0,0000	2,8099	0,0053
NGEL	-0,0034	0,0015	-2,3147	0,0213

Table A5.4: State space univariate model output, for the number of injury accidents in France, for the period 1975-2000

Error variances
 level error variance is 0.000217031
 slope error variance is 1.69763e-011
 seasonal error variance is 2.07209e-006
 irregular variance is 0.000959392

Regression results

	coefficient	standard error	t-value	p-value
LCarbu	0,1377	0,0451	3,0519	0,0025
TE	0,0007	0,0002	3,4245	0,0007
TH	0,0012	0,0003	4,0529	0,0001
HPLUI	0,0000	0,0000	2,9579	0,0033
NGEL	-0,0036	0,0015	-2,4283	0,0157

Table A5.5: State space univariate model output, for the number of injury accidents in France, for the period 1975-2000 - Model with an exposure measure (oil sales).

Motorways and main roads in France

Numerical estimations for the model of the log number of injury accidents, for the period 1975-2000

Error variances

level error variance is 0.00065388 0.00047214 0.00047214 0.00042186	slope error variance is 1.8511e-009 7.2550e-010 7.2550e-010 2.8433e-010
seasonal error variance is 0.00043964 0.00019858 0.00019858 8.9695e-005	irregular variance is 0.0083663 8.1449e-005 8.1449e-005 0.0022372

Regression results

	coefficient	standard error	t-value	p-value
TE	0,0021	0,0006	3,3626	0,0009
TE	0,0002	0,0003	0,7035	0,4823
TH	0,0008	0,0008	0,9826	0,3266
TH	0,0004	0,0004	0,9225	0,3570
HPLUI	0,0001	0,0000	2,6599	0,0082
HPLUI	0,0001	0,0000	4,2527	0,0000
NGEL	0,0053	0,0043	1,2477	0,2131
NGEL	-0,0028	0,0023	-1,1997	0,2312

Table A5.6: State space bivariate model output, for the number of injury accidents on motorways and main roads in France, for the period 1975-2000

Error variances

level error variance is 0.00045782 0.00033783 0.00033783 0.00029454	slope error variance is 1.4947e-008 -1.4586e-008 -1.4586e-008 1.4233e-008
seasonal error variance is 4.7888e-014 -1.5023e-014 -1.5023e-014 1.1304e-005	irregular variance is 0.0081593 8.7591e-005 8.7591e-005 0.0023825

Regression results

	Coefficient	standard error	t-value	p-value
LPaaC	1,0375	0,1102	9,4112	0,0000
	0,0000	0,0000	,NaN	1,0000
	0,0000	0,0000	,NaN	1,0000
LPaRN	0,6862	0,1080	6,3523	0,0000
TE	0,0019	0,0006	3,3040	0,0011
TE	0,0002	0,0003	0,7098	0,4784
TH	0,0005	0,0008	0,6801	0,4969
TH	0,0000	0,0004	0,1135	0,9097
HPLUI	0,0001	0,0000	3,9963	0,0001
HPLUI	0,0001	0,0000	5,0686	0,0000
NGEL	0,0085	0,0040	2,1599	0,0316
NGEL	-0,0020	0,0022	-0,8677	0,3862

Table A5.7: State space bivariate model output, for the number of injury accidents on motorways and main roads in France, for the period 1975-2000 – Models with an exposure measure (vehicle-kilometres)

The whole of the Netherlands

Numerical estimations for the model of the log number of injury accidents, for the period 1987-2005

Error variances
 level error variance is 0.000197427
 slope error variance is 1.70401e-015
 seasonal error variance is 2.62209e-016
 irregular variance is 0.00240132

Regression results

	coefficient	standard error	t-value	p-value
TE	0,0024	0,0003	7,3558	0,0000
TH	0,0017	0,0004	4,5672	0,0000
HPLUI	0,0000	0,0000	1,5553	0,1219
NGEL	-0,0062	0,0032	-1,9477	0,0527

Table A5.8: State space univariate model outputs, for the number of injury accidents in the Netherlands, for the period 1987-2005.

Error variances
 level error variance is 0.000213724
 slope error variance is 7.29565e-018
 seasonal error variance is 5.13812e-013
 irregular variance is 0.002339

	coefficient	standard error	t-value	p-value
LDriverkm	0,0231	0,0126	1,8394	0,0672
TE	0,0023	0,0003	7,0359	0,0000
TH	0,0016	0,0004	4,0956	0,0001
HPLUI	0,0000	0,0000	1,8926	0,0597
NGEL	-0,0064	0,0032	-2,0348	0,0430

Table A5.9: State space univariate model outputs, for the number of injury accidents in the Netherlands, for the period 1987-2005. Model with an exposure measure (driver-kilometers)

Motorways and rural roads in the Netherlands

Numerical estimations for the model of the log number of injury accidents, for the period 1987-2005

Error variances

level error variance is 0.0035696 0.0020047 0.0020047 0.0011653	slope error variance is 2.2555e-007 2.8265e-007 2.8265e-007 3.5420e-007
seasonal error variance is 0.0010064 0.00023946 0.00023946 5.6976e-005	irregular variance is 2.7483e-013 6.7131e-010 6.7131e-010 0.0053410

Regression results

	coefficient	standard error	t-value	p-value
TE	0,0016	0,0004	3,8714	0,0001
TE	0,0028	0,0005	5,5988	0,0000
TH	0,0003	0,0005	0,6150	0,5392
TH	0,0012	0,0006	2,0961	0,0372
HPLUI	0,0001	0,0000	5,2881	0,0000
HPLUI	0,0000	0,0000	1,0224	0,3077
NGEL	0,0005	0,0044	0,1035	0,9176
NGEL	-0,0013	0,0049	-0,2708	0,7868

Table A5.10.: State space bivariate model outputs, for the numbers of injury accidents on the motorways and rural roads in the Netherlands, for the period 1987-2005

Error variances

level error variance is 0.0088016 0.0037643 0.0037643 0.0016100	slope error variance is 3.1309e-007 3.8297e-007 3.8297e-007 4.6846e-007
seasonal error variance is 1.4337e-013 -1.1551e-012 -1.1551e-012 1.0275e-010	irregular variance is 9.9296e-012 1.1714e-010 1.1714e-010 0.0057713

Regression results

	coefficient	standard error	t-value	p-value
TE	0,0020	0,0004	4,5959	0,0000
TE	0,0025	0,0005	5,0167	0,0000
TH	-0,0002	0,0005	-0,4928	0,6226
TH	0,0009	0,0006	1,5853	0,1143
HPLUI	0,0001	0,0000	5,9321	0,0000
HPLUI	0,0000	0,0000	1,9508	0,0523
NGEL	-0,0043	0,0044	-0,9680	0,3341
NGEL	-0,0006	0,0050	-0,1295	0,8971

Table A5.11: State space bivariate model outputs, for the numbers of injury accidents with motorvehicles on the motorways and rural roads in the Netherlands, for the period 1987-2005.

The Athens region

Numerical estimations for the model of the log number of injury accidents, for the period 1987-2005 and 1997-2005

Error variances
 level error variance is 0.00128914
 slope error variance is 3.71242e-016
 seasonal error variance is 3.63635e-012
 irregular variance is 0.00263571

Regression results

	coefficient	standard error	t-value	p-value
LTotalveh				
TE	0,0003	0,0006	0,4794	0,6321
TH	0,0021	0,0004	5,1465	0,0000
HPLUI	-0,0005	0,0001	-3,7656	0,0002

Table A5.12: State space bivariate model outputs, for the numbers of injury accidents in the region of Athens, for the period 1987-2005.

Error variances
 level error variance is 4.87845e-016
 slope error variance is 6.91738e-006
 seasonal error variance is 3.5027e-012
 irregular variance is 0.00426146

Regression results

	coefficient	standard error	t-value	p-value
LTotalveh	0,1148	0,1117	1,0284	0,3061
TE	-0,0001	0,0010	-0,1231	0,9023
TH	0,0031	0,0006	4,7516	0,0000
HPLUI	-0,0003	0,0002	-1,7093	0,0904

*Table A5.13: State space bivariate model outputs, for the numbers of injury accidents in the region of Athens, for the period 1997-2005. Models **with an exposure measure (vehicles)**.*