Test modelling single accidents with the basic evaluation model

Frits Bijleveld & Jacques Commandeur

D-2006-3

Test modelling single accidents with the basic evaluation model

D-2006-3
Frits Bijleveld & Jacques Commandeur
Leidschendam, 2006
SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number:	D-2006-3
Title:	Test modelling single accidents with the basic evaluation model
Author(s):	Frits Bijleveld & Jacques Commandeur
Project leader:	Paul Wesemann
Project number SWOV:	40.103
Keywords:	Mathematical model, fatality, vehicle mile, trend (stat), development, forecast, safety, traffic, statistics, unobserved component, state space method, explanatory variable, smoothing, Netherlands.
Contents of the project:	This report elaborates and illustrates the basic evaluation model for the evaluation and exploration of developments in Dutch road safety. Two methods are presented for including explanatory variables in the model.
Number of pages:	30 + 7
Price:	€ 10,-
Published by:	SWOV, Leidschendam, 2006

SWOV Institute for Road Safety Research
P.O. Box 1090
2260 BB Leidschendam
The Netherlands
Telephone +31 70 317 33 33
Telefax +31 70 320 12 61
E-mail info@swov.nl
Internet www.swov.nl

Summary

This report elaborates and illustrates the proposed methodology for the evaluation and exploration of developments in Dutch road safety in the SWOV projects ‘Road safety assessments’ (Verkeersveiligheidsbalansen) and ‘Road safety outlooks’ (Verkeersveiligheidsverkenningen).
Previous work in this area resulted in what is called the basic evaluation model, which considers the development of traffic safety to be the product of the developments in two latent, unobserved variables: exposure and risk. The basic evaluation model is a bivariate local linear trend model, and well-suited for handling the dependencies in time series data, thus yielding residuals satisfying the model assumption of independence.
In the present report the basic evaluation model is extended to the incorporation of explanatory variables. Two methods are presented for including explanatory variables in the basic evaluation model. The first method uses a standard regression set-up in which the explanatory variables are treated as fixed and known. In the second method, on the other hand, they are treated as being subject to stochastic variation. The two methods are applied to evaluate the effects of wet weather conditions and drink-driving on the annual numbers of single motor vehicle accidents involving people being killed or severely injured in the Netherlands. It is found that only the second method uncovers the expected relationships between the explanatory variables and the risk of getting killed or severely injured in single motor vehicle accidents.

Contents

1. Introduction

This report elaborates and illustrates the proposed methodology for the evaluation of developments in Dutch road safety in the SWOV projects ‘Traffic safety evaluations’ (Verkeersveiligheidsbalansen) and ‘Traffic safety explorations ’ (Verkeersveiligheidsverkenningen), as presented earlier in Bijleveld (1999), Bijleveld and Commandeur (2004), and Gould, Bijleveld and Commandeur (2004). The objective of the first mentioned project is to obtain explanations for disaggregated developments in road traffic safety in the Netherlands, while that of the second project is to make forecasts for future developments of Dutch road safety, based on the modelled disaggregated developments in the past.

As discussed in Bijleveld and Commandeur (2004), a central issue in the modelling of developments over time is that the observations always consist of repeated measurements of one and the same phenomenon, like for example the annual number of road traffic fatalities as recorded over a number of years. Although the individual fatalities are independent of one another over the years, their annual sums are not, being the result of a traffic process that slowly evolves over time. The number of road traffic fatalities observed in a certain year is therefore often quite a good predictor for the number of fatalities that will occur in the following year. Since the observations in such time series are not independent of one another, special care must be taken in dealing with the dependencies in the data. Failing to do so usually results in serially correlated model residuals, and therefore in flawed confidence limits and significance tests.

This was one of the reasons to use a family of techniques known as structural time series or unobserved component models by state space methods (see Harvey, 1989; Durbin and Koopman, 2001), since these are capable of explicitly handling the dependencies between consecutive observations in time series data. A dedicated multivariate state space model called the basic evaluation model was developed, in which the development of traffic safety is assumed to be the product of the developments of two latent, unobserved variables: exposure and risk. Special features of the latter model are that the observed mobility and road safety figures are both assumed to be subject to measurement error, and that the risk is completely treated as a latent variable. The reason that the risk is completely modelled as a latent variable is that this is indeed what risk is: an unobservable entity that can not be measured directly. Although this is not discussed in the present report, the basic evaluation model can also be used to obtain estimates for future developments of road safety.

In this report, the methodology discussed in Bijleveld and Commandeur (2004) is extended to include explanatory variables. Specifically, the basic evaluation model is applied to the development of single motor vehicle accidents involving people being killed or severely injured in the Netherlands in the years 1985-2003. A single accident is an accident in which no other traffic participant is involved. These data and other variables used in the analyses are discussed in *Chapter 2*. The basic evaluation model is discussed in *Chapter 3*, and the results of its application to single motor vehicle accidents are shown in *Chapter 4*.

Then, two explanatory variables are added to the basic evaluation model: the annual proportion of time with wet weather, and (an estimate of) the annual proportion of car drivers circulating in traffic with a BAC of more than 0.05 percent, and the effects of these two variables on the risk of being involved in a single motor vehicle accident is estimated. These two variables are used for the simple and pragmatic reason that at the time of writing at least some information was available about these variables for the period 1985-2003.

In *Chapters 5 and 6*, two methods are presented and illustrated for incorporating explanatory variables in the basic evaluation model. The first method uses a standard regression set-up in which the explanatory variables are treated as fixed and known. In the second method, on the other hand, they are treated as being subject to stochastic variation. The results of the two methods are compared and their relative merits are discussed in *Chapter 7*. In this comparison special consideration is given to the plausibility of the estimated effects of the explanatory variables on the risk of getting involved in single motor vehicle accidents with fatal or serious consequences.

2. The data

The data used in this report are of the Dutch annual figures of the total number of single motor vehicle accidents involving people being Killed or Severely Injured (KSI), and the total number of car kilometres travelled (travel kilometres) in the period 1985-2003 (see columns two and three in *Table 1*). Note that the travel kilometres for 2003 are missing.

year	single accidents KSI	travel kilometres for cars	single accidents KSI with wet weather	single drink driving accidents KSI	proportion of time with wet weather	percentage of drink driving in weekend nights
1985	1408	60.6185	449	540	0.06817	missing
1986	1442	63.2202	363	586	0.06773	missing
1987	1355	64.7728	383	405	0.08147	8.0
1988	1315	70.6459	382	382	0.08471	6.0
1989	1334	71.2214	313	437	0.06144	6.0
1990	1310	73.5699	323	387	0.06035	missing
1991	1271	72.1167	294	384	0.05716	3.9
1992	1195	76.3503	310	297	0.08322	4.0
1993	1145	73.7701	322	317	0.09535	4.2
1994	1141	73.6165	331	333	0.09855	4.9
1995	1296	77.4663	330	370	0.07586	4.7
1996	1297	78.7220	358	383	0.06227	4.4
1997	1243	80.2868	293	370	0.06377	4.3
1998	1175	82.1500	358	326	0.10145	4.5
1999	1275	85.2666	323	386	0.07279	4.3
2000	1269	85.9485	324	322	0.08330	4.6
2001	1194	86.9959	347	281	0.09128	4.2
2002	1172	88.7302	264	297	0.08285	missing
2003	1184	missing	226	330	0.05718	missing

Table 2.1*. The data*

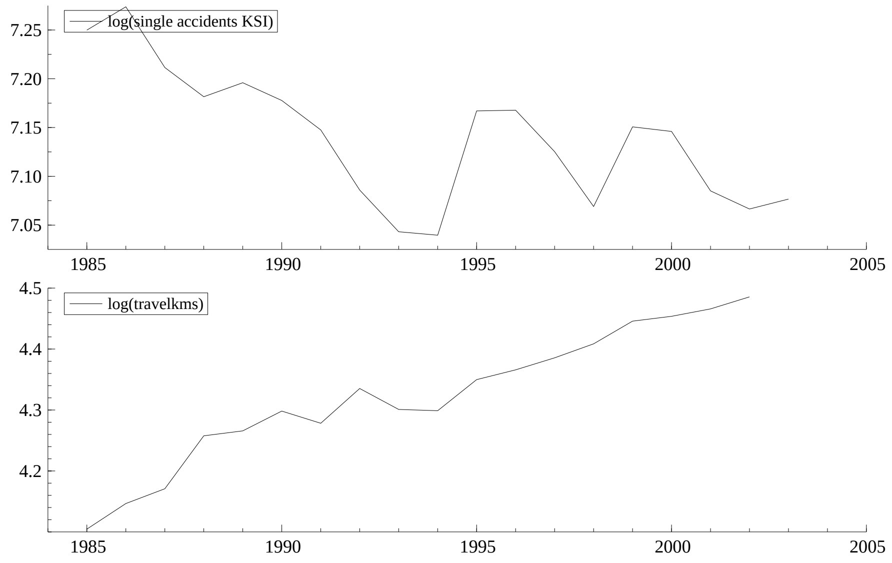

Figure 2.1. *Logarithm of single accidents KSI (top) and of car kilometres travelled (bottom)*.

Plots of the logarithm of the number of single accidents KSI, and of the logarithm of the number of travel kilometres, are shown in *Figure 2.1*. The reason why their logarithms are taken will become clear in *Chapter 3*. The remaining variables in *Table 1* will be explained in *Chapters 5* and *6*.

3. The basic evaluation model (BEM)

At the core of all the models discussed in the present report is the following multiplicative model:

$$\text{traffic safety} = \text{exposure} \times \text{risk} \times e^{\text{error}}, \qquad (1)$$

where 'traffic safety' – further denoted by F_t – is the dependent variable, and 'exposure' and 'risk' are the independent variables.
It is assumed that both 'exposure' and 'risk' are latent unobserved variables, and it is further assumed that the observed development in mobility – as measured in travel kilometres, and further denoted by M_t – acts as an indicator variable for the unobserved 'exposure':

$$M_t = \text{exposure}_t \times e^{\text{error}}. \qquad (2)$$

It follows from (2) that the observed mobility figures are assumed to be contaminated with measurement error, and are not identical to, but only proportional with the unobserved variable 'exposure'.
Taking the logarithms of (2) and (1), respectively, the following additive bivariate log-linear model is obtained:

$$\log(M_t) = \log(\text{exposure}_t) + \text{error}_t^{(1)}, \qquad (3)$$

and

$$\log(F_t) = \log(\text{exposure}_t) + \log(\text{risk}_t) + \text{error}_t^{(2)}, \qquad (4)$$

for $t = 1, \ldots, n$, where n is the number of observations (i.e., time points). Note that the observed travel kilometres (M_t) are treated as a *dependent* variable in (3), but that the unobserved exposure derived from (3) is treated as an *independent* variable in (4).
The observed figures in F_t are not independent observations, due to the fact that they involve repeated measurements of one and the same phenomenon in time. The same applies to the variable M_t. To accommodate for these dependencies, the unobserved independent variables 'exposure' and 'risk' are modelled using two so-called local linear trend models (see, e.g., Durbin & Koopman, 2001):

$$\text{trend}[\log(\text{exposure}_{t+1})] = \text{level}[\log(\text{exposure}_t)] + \text{slope}[\log(\text{exposure}_t)] + \text{error}_t^{(3)} \qquad (5)$$

$$\text{slope}[\log(\text{exposure}_{t+1})] = \text{slope}[\log(\text{exposure}_t)] + \text{error}_t^{(4)} \qquad (6)$$

$$\text{trend}[\log(\text{risk}_{t+1})] = \text{level}[\log(\text{risk}_t)] + \text{slope}[\log(\text{risk}_t)] + \text{error}_t^{(5)} \qquad (7)$$

$$\text{slope}[\log(\text{risk}_{t+1})] = \text{slope}[\log(\text{risk}_t)] + \text{error}_t^{(6)}. \qquad (8)$$

In these four equations, the level and the slope terms are similar to the well-known intercept and regression weight of classical linear regression, respectively. The crucial difference is that the level and slope terms in (5) to (8) are allowed to *change over time*, whereas they are constant in classical regression.
In fact, it is not very difficult to prove that when all error terms in (5) to (8) are restricted to be equal to zero, equations (5) and (7) can be written as

$$\text{trend}[\log(\text{exposure}_t)] = a + bt + \text{error}_t^{(3)},$$

and

$$\text{trend}[\log(\text{risk}_t)] = c + dt + \text{error}_t^{(5)},$$

where $t = 1, \ldots, n$ is time, a and c are intercepts, and b and d are regression weights. This means that the local linear trend model then collapses to a classical linear regression model, and that linear regression is just a special case of the state space methods used in this report.

Together, equations (3) through (8) define the so-called basic evaluation model (BEM), as presented in Bijleveld (1999), Bijleveld and Commandeur (2004), and Gould, Bijleveld and Commandeur (2004). For a more technical discussion of the BEM, we refer to *Appendix 1*. In the next chapter the BEM is applied to the single accidents KSI and the travel kilometres shown in *Table 1*.
All analyses in the present report were performed in Ox/SsfPack (see Doornik, 2002; Koopman, Shepard & Doornik, 1999) using the univariate approach to multivariate state space methods in which the observation disturbances are estimated by putting them in the state vector (see Durbin & Koopman, 2001, p.131).

4. Single accidents

Applying the basic evaluation model defined in (3) through (8) to the logarithm of the Dutch annual numbers of single motor vehicle accidents KSI and of the number of travel kilometres given in *Table 1* yields the following results.
One hundred random starts were used to ensure convergence to the global maximum of the log-likelihood function. Out of these one hundred random starts, the so-called BFGS (Broyden-Fletcher-Goldfarb-Shannon) algorithm converged fourteen times to the best log-likelihood value of 54.9339. The value of the Akaike Information Criterion (AIC) for this model equals -2.1018.
The estimated variance matrix for the level disturbances in (5) and (7) (the first block in block-diagonal matrix Q_t defined in *Appendix 1* is

$$\begin{bmatrix} 0.000067 & 0.000339 \\ 0.000339 & 0.001720 \end{bmatrix},$$

while the estimated variance matrix for the slope disturbances in (6) and (8) (the second block in block-diagonal matrix Q_t defined in *Appendix 1* equals

$$\begin{bmatrix} 0.000076 & -0.000153 \\ -0.000153 & 0.000308 \end{bmatrix}.$$

The estimated variance matrix for the irregular components in (3) and (4) (matrix H_t defined in *Appendix 1* equals

$$\begin{bmatrix} 0.000280 & 0.000008 \\ 0.000008 & 0.0000003 \end{bmatrix}.$$

Plots of the results of the analysis are displayed in *Figures 4.1* to *4.3*. The top graph in *Figure 4.1* shows the log of the observed Dutch annual travel kilometres, together with the trend in the unobserved exposure (corresponding to equation (5)), including its 95% confidence interval. The middle graph displays the development of the slope component for the exposure (equation (6)), including its 95% confidence interval, while the bottom graph contains the irregular component of equation (3).
Note that the confidence intervals for the trend and slope in exposure are larger in 2003, reflecting the fact that no information on travel kilometres for cars is available for that year.

The top graph in *Figure 4.2* shows the trend for the unobserved risk (corresponding to equation (7)), including its 95% confidence interval, while the bottom graph displays the development of the slope component for the risk (equation (8)), including its 95% confidence interval.

Figure 4.3 finally contains both the log of the observed Dutch single accidents KSI as well as the modelled development, including its 95% confidence interval. The modelled development is the sum of the trend in exposure (as shown in *Figure 4.1*) and the trend in risk (as shown in *Figure 4.2*). The bottom graph contains the irregular component of equation (4).

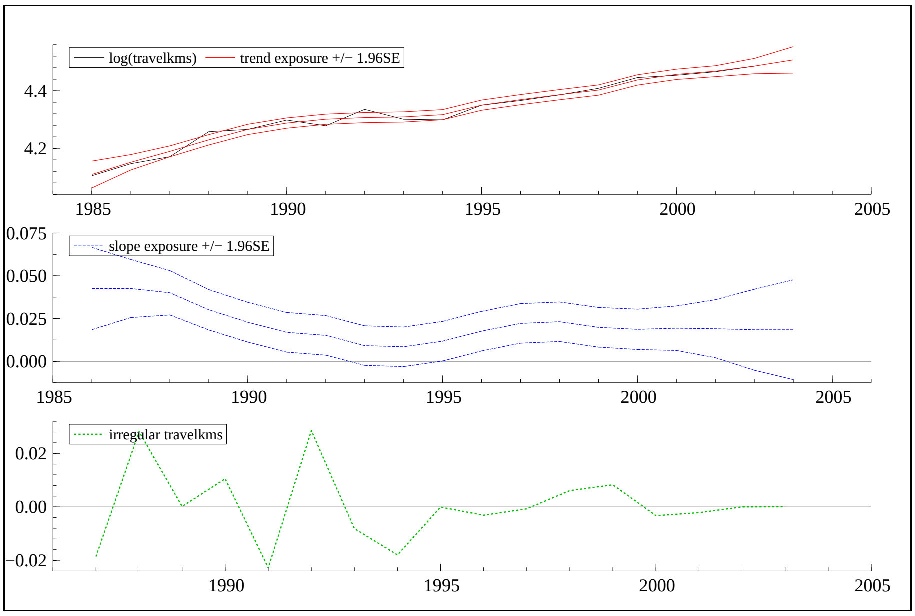

Figure 4.1. *Dutch travel kilometres and exposure in the years 1985-2003: trend, slope (including 95% confidence intervals), and irregular.*

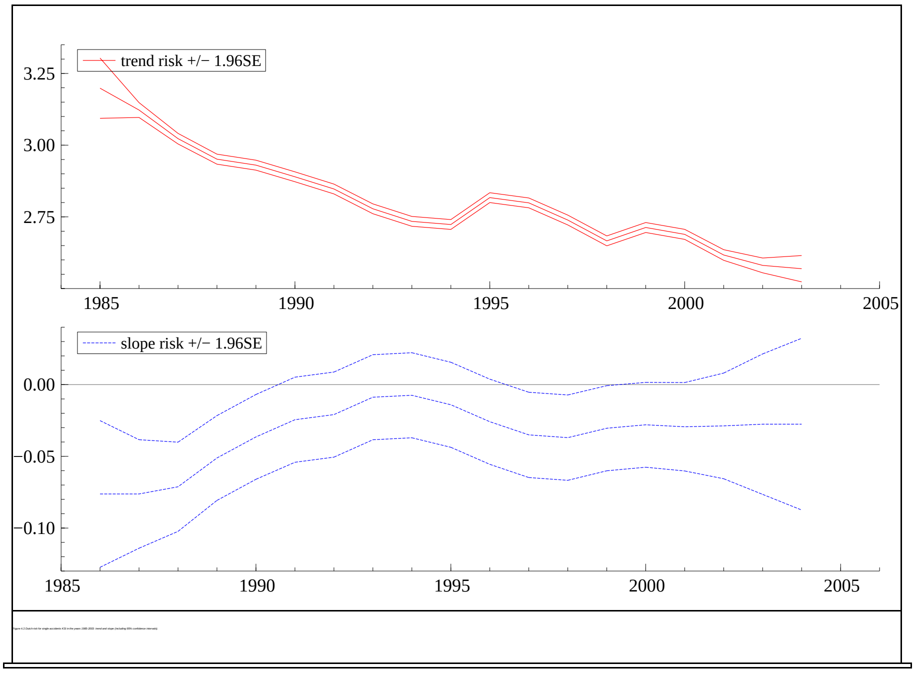

Figure 4.2. *Dutch risk for single accidents KSI in the years 1985-2003: trend and slope (including 95% confidence intervals)*

Figure 4.3. *Dutch development in single accidents KSI in the years 1985-2003 (including its 95% confidence interval), and irregular.*

Note that the disturbances of the irregular components in *Figures 4.1* and *4.3* are (much) larger in the first part of the series than in the second part. This heteroscedasticity in the residuals suggests that the observed mobility figures are more contaminated with measurement error in earlier years than in more recent years.

In the next chapters, two methods are presented for adding explanatory variables to the basic evaluation model discussed so far.

5. Extending the BEM with explanatory variables: first method

In this and in the following chapter, two explanatory variables are considered for inclusion in the basic evaluation model. The two variables were chosen for the simple reasons that (some) information on them was available for the period under investigation, and that they are known to have an effect on road safety.

The first explanatory variable is the annual *proportion of time with precipitation* (rain, hail, snow), as measured in De Bilt in the years 1985-2003. In the sequel, this variable will loosely be denoted as the *proportion of time with wet weather*. The observations for this variable are given in the sixth column of *Table 2.1*, and a plot of the logarithm of the variable is shown at the top of *Figure 5.1*.

The second explanatory variable is the annual *proportion of car drivers circulating in traffic with a BAC (Blood Alcohol Content) of more than 0.05 percent*. These yearly proportions of drink driving in Dutch road traffic were estimated by consulting two different data sources.

The first source, SWOV/AVV Drink and driving habits (SWOV/AVV Rij- en drinkgewoonten) is a time series of annual percentages of car drivers with a BAC larger than 0.05 percent. Unfortunately, these percentages refer to *weekend nights only* (that is, to Friday to Sunday nights from 22.00 PM to 04.00 AM). For the years 1985 to 2003, these percentages are shown in the last column in *Table 2.1*. Note that the percentages are missing for the years 1985, 1986, 1990, 2002 and 2003.

The second source of information is a study performed in the years 2000 to 2004 in Tilburg, a city located in the south of the Netherlands, where percentages of drink driving with a BAC larger than 0.05 percent were obtained for car, van, and minibus drivers, *both* for weekend nights *and* for the entire week, 7x24hrs (Mathijssen & Houwing, 2005). The latter percentages are 4.8 and 1.1, respectively, implying a ratio of 4.8/1.1 = 4.4 between drink driving percentages in weekend nights and for the entire week.

On the rather strong assumption that this ratio can be generalized to the rest of the Netherlands, an estimate of the percentage of drink driving in the whole of the Netherlands during the entire week was obtained by dividing the above mentioned percentages from ‘AVV/SWOV Drink and driving habits’ for weekend nights by a factor of 4.4. The thus estimated percentages were finally transformed into proportions by dividing the percentages by 100. A graph of the logarithm of the thus estimated proportions of drink driving for the entire week (and for the whole of the Netherlands) is shown as a dashed line at the bottom of *Figure 5.1*.

We finally note that the first method for evaluating the effects of explanatory variables (to be discussed below) resulted in numerical problems when an explanatory variable contained missing observations. For the first method, an ad hoc method was therefore used to obtain estimates of the unknown percentages of drink driving during weekend nights in the years 1985, 1986, 1990, 2002, and 2003 (see *Table 2.1*). Since the last known percentage is

12% for 1983, the percentages in 1985 and 1986 were linearly interpolated from the latter percentage, and that for 1987 (8%), yielding estimates of 10% for 1985, and of 9% for 1986. The same procedure was applied to 1989 (6%) and 1991 (3.9%), yielding an estimated percentage of 5% for 1990. The percentages for 2002 and 2003 were assumed to be equal to the percentage of 2001, which is 4.2%. Just as before, these percentages were divided by ten (to get proportions), and then by 4.4 (yielding estimated proportions of drink driving for the whole week in these years).

In the first method for introducing explanatory variables in the basic evaluation model, the latter variables are simply added to risk component (7) discussed in Chapter 3. For one explanatory variable x_t, this yields

$$\text{trend}[\log(\text{risk}_{t+1})] = \text{level}[\log(\text{risk}_t)] + \text{slope}[\log(\text{risk}_t)] + \beta\Delta x_t + \text{error}_t^{(5)}, \quad (9)$$

where β is an unknown regression weight, and $\Delta x_t = x_{t+1} - x_t$ for $t = 1, \ldots, n$. It may be noted that the addition of an explanatory variable to the risk component gives the same result as adding the explanatory variable itself to (4), that is,

$$\log(F_t) = \log(\text{exposure}_t) + \log(\text{risk}_t) + \beta x_t + \text{error}_t^{(2)}.$$

(see, Harvey, 1989; Commandeur, 2004). A more technical discussion of the model is given in *Appendix 2*.

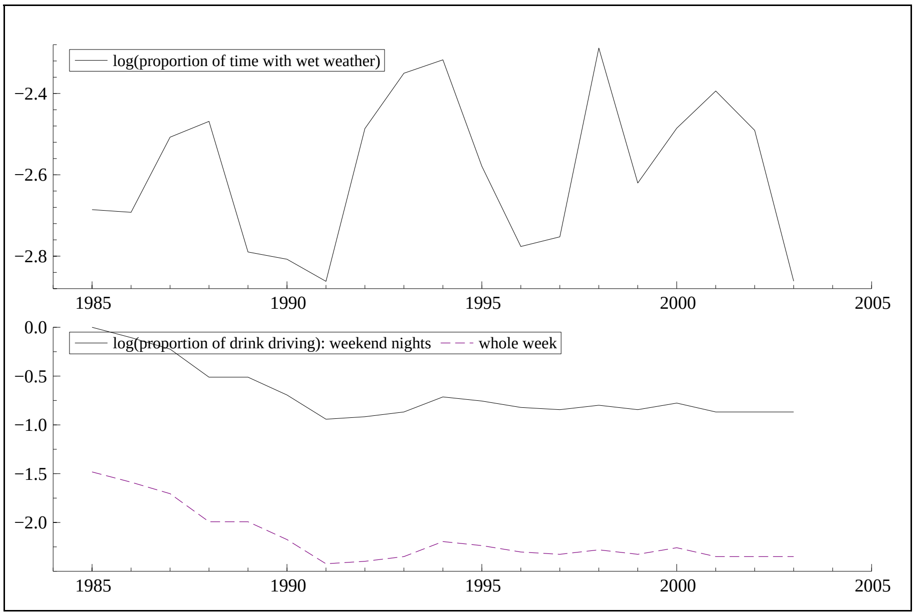

Figure 5.1. *Logarithm of proportion of time with wet weather (top), and of proportion of drink driving during weekend nights and the whole week (bottom)*.

Adding the logarithm of the ‘proportion of time with wet weather’, and of the estimated ‘proportion of drink driving’ to equation (9) gives the following results.

One hundred random starts were used to ensure convergence to the global maximum of the log-likelihood function. Out of these one hundred random starts, the BFGS algorithm converged four times to the best log-likelihood value of 59.0896. The value of the Akaike Information Criterion equals -2.2152, which is an improvement upon the BEM without explanatory variables (see *Chapter 4*).

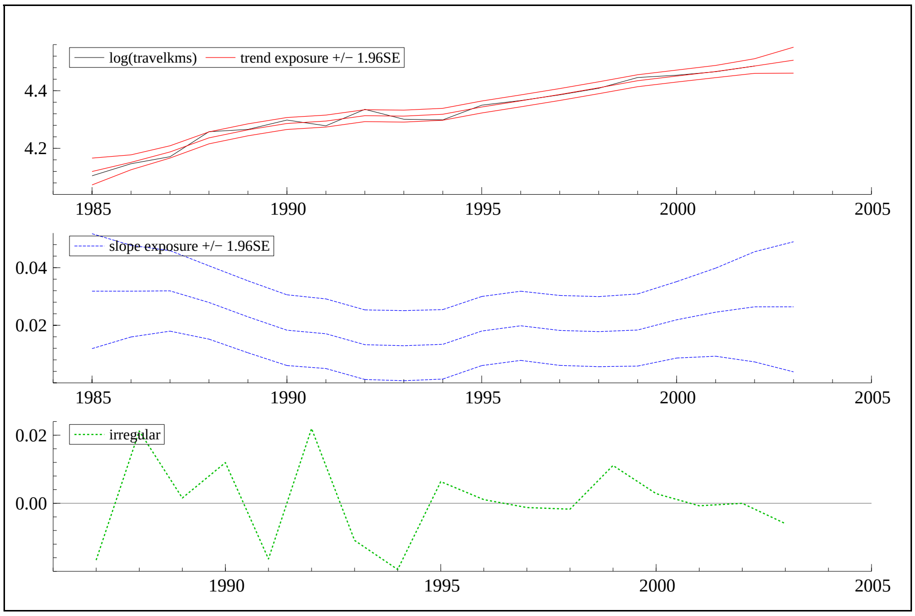

Figure 5.2*. Dutch exposure in the years 1985-2003: trend, slope, and irregular.*

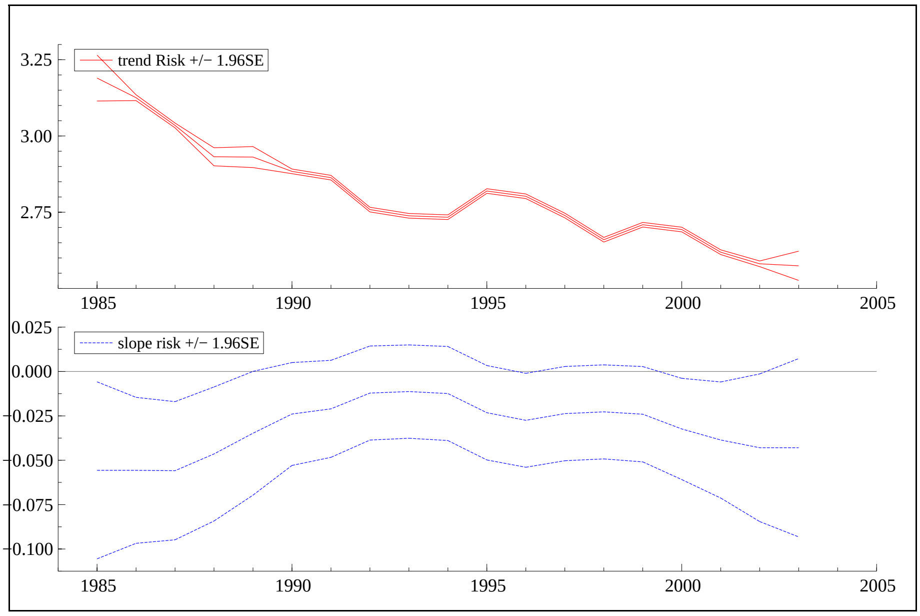

Figure 5.3*. Dutch risk for single accidents KSI including wet weather in the years 1985-2003: trend and slope.*

Plots of the results of the analysis are displayed in *Figures 5.2* to *5.4*. The top graph in *Figure 5.2* shows the log of the Dutch annual travel kilometres,

together with the trend in exposure, including its 95% confidence interval. The middle graph displays the slope component for the exposure, including its 95% confidence interval, while the bottom graph contains the irregular component of equation (3).
Again the confidence intervals for the trend and slope in exposure are larger in 2003, because no information on travel kilometres for cars is available for that year.
The top graph in *Figure 5.3* shows the modelled development in risk (corresponding to equation (9)), including its 95% confidence interval. The bottom graph displays the slope component for the risk, including its 95% confidence interval.

Figure 5.4*. Dutch development in single accidents KSI in the years 1985-2003), and irregular.*

Figure 5.4 again shows both the observed single accidents KSI and their modelled development, including the 95% confidence interval (top), and the bottom graph contains the irregular component of equation (4).

The value of the regression weight in (9) for the explanatory variable ‘log(proportion of time with wet weather)’ equals –0.17, and the *t*-test for the regression weight is significant (t = 5.35). The value of the regression weight in (9) for the explanatory variable ‘log(proportion of drink driving)’ equals 0.09, and the *t*-test for this regression weight is not significant (t = 1.32).

Since the effect of the explanatory variable ‘proportion of drink driving’ is not significant, the analysis was repeated without this variable. Out of one hundred random starts, the algorithm converged seven times to the best log-likelihood value of 60.1896. The value of the Akaike Information Criterion for this model equals -2.3258, which is an improvement upon the BEM including both explanatory variables (see the previous analysis).
The estimated variance matrix for the level disturbances in (5) and (9) (matrix *B* in *Appendix 2*) is

$$\begin{bmatrix} 0.000136 & 0.000024 \\ 0.000024 & 0.000602 \end{bmatrix},$$

while the estimated variance matrix for the slope disturbances in (6) and (8) (matrix *C* in *Appendix 2*) equals

$$\begin{bmatrix} 0.000055 & -0.000157 \\ -0.000157 & 0.000449 \end{bmatrix}.$$

The estimated variance matrix for the observation disturbances in (3) and (4) (matrix H_t in *Appendix 2*) is

$$\begin{bmatrix} 0.000280 & 0.000008 \\ 0.000008 & 0.0000002 \end{bmatrix}.$$

Graphs of the results of this analysis are not shown because they are virtually identical to the ones presented in *Figures 5.2* to *5.4*.
Compared to the previous analysis, the value of the regression weight in (9) for the explanatory variable 'log(proportion of time with wet weather)' is unchanged; it still equals -0.17 (t = 5.35).

Unfortunately, in both analyses the value of the regression weight for 'proportion of time with wet weather' is not in the expected direction. Since this explanatory variable and the dependent variable single accidents KSI are both analysed in their logarithms, the regression weight can be interpreted as a so-called *elasticity*. A 1% increase in the proportion of time with wet weather is therefore associated with a 0.17% *decrease* in single accidents KSI. The negative sign of the regression weight also indicates that larger proportions of time with wet weather are associated with smaller risks of becoming involved in single accidents KSI.
It is also strange that no effect is found for drink driving on the number of single accidents KSI (although the sign of the corresponding regression weight is at least in the expected direction).
In the next chapter an alternative and new approach for adding explanatory variables to the BEM is presented, which does yield the expected results.

6. Extending the BEM with explanatory variables: second method

The second method for adding explanatory variables to the basic evaluation model is completely new, and consists of a multivariate time series model with the following *six* dependent variables:

$$y_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \\ y_t^{(6)} \end{pmatrix} = \begin{pmatrix} \log(p_t^{wet}) \\ \log(p_t^{alc}) \\ \log(M_t) \\ \log(F_t) \\ \log(F_t^{wet}) \\ \log(F_t^{alc}) \end{pmatrix},$$

where

- p_t^{wet} is the observed proportion of time in the year with precipitation (rain, hail, snow, as measured in De Bilt) at time point *t*,
- p_t^{alc} is the observed proportion of drink driving at time point *t*,
- M_t is the observed number of kilometres travelled by cars at time point *t*,
- F_t is the observed number of single motor vehicle accidents involving people being Killed or Severely Injured (KSI) at time point *t*,
- F_t^{wet} is the observed number of single motor vehicle accidents KSI that occurred during precipitation (rain, hail, snow) (see the fourth column in *Table 2.1*) at time point *t*,
- F_t^{alc} is the observed number of single motor vehicle drink driving accidents at time point *t* (see the fifth column of *Table 2.1*).

The starting point of the method is the idea that the total (unobserved) exposure in the BEM can be divided in two additive parts: one part that is the exposure in wet weather conditions, and a second part that is the exposure in dry weather conditions. Let $0 < f_t^{wet} < 1$ denote the *proportion* of the total exposure in wet weather conditions, then it is always true that

$$\begin{aligned} \text{exposure}_t &= \text{exposure}_t(\text{in wet weather}) + \text{exposure}_t(\text{in dry weather}) \\ &= (\text{exposure}_t \times f_t^{wet}) + (\text{exposure}_t \times (1 - f_t^{wet})). \end{aligned}$$

The next step is to see that the observed explanatory variable p_t^{wet} can be considered as an appropriate indicator variable for this unobserved proportion f_t^{wet} . Therefore, the following equation is added to equations (3) and (4) of the BEM:

$$\log(p_t^{wet}) = \log(f_t^{wet}) + \mathrm{error}_t . \qquad (10)$$

Just as the observed number of travel kilometres is used in the BEM as an indicator for the unobserved exposure, so is the observed proportion p_t^{wet} used as an indicator for the unobserved proportion f_t^{wet} .
Next, the observed number of single motor vehicle accidents KSI that occurred in wet circumstances (F_t^{wet}) can be modelled as the product of two unobserved components: the exposure in wet weather conditions and the *risk* in wet weather conditions, as follows:

$$F_t^{wet} = \text{exposure}_t\text{(in wet weather)} \times \text{risk}_t\text{(in wet weather)} \times e^{\text{error}},$$
$$= (\text{exposure}_t \times f_t^{wet}) \times (\text{risk}_t \times k_t^{wet}) \times e^{\text{error}},$$

where k_t^{wet} is an unknown constant. The value of k_t^{wet} determines how much the total risk of becoming involved in single accidents KSI needs to be in- or decreased in order to explain the observed number of single accidents KSI *that occurred in wet weather conditions*. Taking the logarithm, we obtain the additive model

$$\log(F_t^{wet}) = \log(\text{exposure}_t) + \log(f_t^{wet}) + \log(\text{risk}_t) + \log(k_t^{wet}) + \mathrm{error}_t \qquad (11).$$

Summarizing so far, the second method for adding an explanatory variable to the BEM consists of equations (3), (4), (10) and (11).
Again note that the observed explanatory variable p_t^{wet} is treated as a *dependent* variable in (10), but that the unobserved f_t^{wet} derived from (10) is handled as an *independent* variable in (11) (just as was done for the number of travel kilometres in *Chapter 3*).
Finally, to equations (5) through (8) of the BEM two equations are added:

$$\log(k_{t+1}^{wet}) = \log(k_t^{wet}), \qquad (12)$$

and

$$\text{level}[\log(f_{t+1}^{wet})] = \text{level}[\log(f_t^{wet})] + \mathrm{error}_t . \qquad (13)$$

The latter equation defines a local level model.

Applying the same procedure to the explanatory variable 'proportion of drink driving' (p_t^{alc}), we obtain the following additional equations:

$$\log(p_t^{alc}) = \log(f_t^{alc}) + \text{error}_t\ , \qquad (14)$$

$$\log(F_t^{alc}) = \log(\text{exposure}_t) + \log(f_t^{alc}) + \log(\text{risk}_t) + \log(k_t^{alc}) + \text{error}_t \qquad (15)$$

$$\log(k_{t+1}^{alc}) = \log(k_t^{alc}), \qquad (16)$$

and

$$\text{level}[\log(f_{t+1}^{alc})] = \text{level}[\log(f_t^{alc})] + \text{error}_t\ . \qquad (17).$$

A more technical (state space) notation for the complete model is given in *Appendix 3*.

The second method for adding explanatory variables to the BEM involves all data in *Table 2.1*. Plots of the logarithm of the observed annual numbers for F_t^{wet} and F_t^{alc} are shown in *Figures 6.6* and *6.7*, respectively. In the following analysis, the estimated proportions of drink driving during the whole week for the years 1985, 1986, 1990, 2002, and 2003 were treated as missing (as can be seen in *Figure 6.2*).
Using one hundred random starts, the best value for the log-likelihood of 94.7362 is obtained in three cases. The hyper parameter estimates are (see *Appendix 3* for their meaning):

$$H_t = \begin{bmatrix} 0.027460 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.019711 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.000243 & 0.000080 & 0.000157 & -0.000419 \\ 0 & 0 & 0.000080 & 0.000467 & -0.001067 & 0.001609 \\ 0 & 0 & 0.000157 & -0.001067 & 0.003020 & -0.004318 \\ 0 & 0 & -0.000419 & 0.001609 & -0.004318 & 0.009445 \end{bmatrix}$$

and

$$Q_t = \begin{bmatrix} 0.005740 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.002941 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.000302 & -0.000157 & 0 & 0 \\ 0 & 0 & -0.000157 & 0.001577 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.000006 & -0.000016 \\ 0 & 0 & 0 & 0 & -0.000016 & 0.000043 \end{bmatrix}$$

Graphs of the observed time series as well as their modelled developments and 95% confidence intervals are shown in *Figures 6.1* to *6.7*.
The modelled development in *Figure 6.1* corresponds to equation (13), together with the irregular component in (10). The modelled development in *Figure 6.2* corresponds to equation (17), together with the irregular

component in (14). The modelled developments in *Figure 6.3* correspond to equations (5) and (6), together with the observation errors in (3); the modelled developments in *Figure 6.4* correspond to equations (7) and (8). The modelled development in *Figure 6.5* corresponds to equation (4), together with its irregular component. Finally, the modelled development in *Figure 6.6* corresponds to equation (11), and that in *Figure 6.7* to equation (15), both shown with their irregular components.

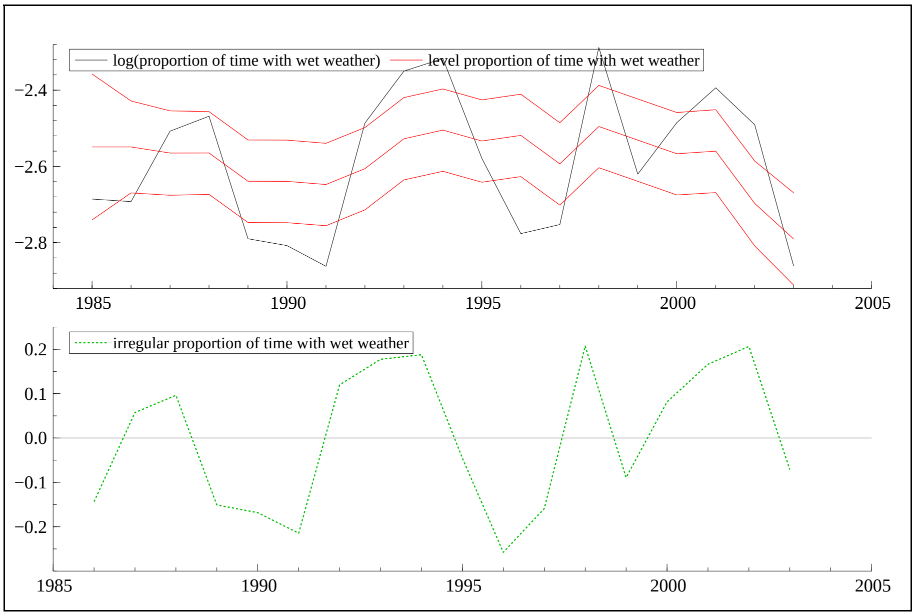

Figure 6.1. *Observed and modelled development of proportion of time with wet weather, and irregular component.*

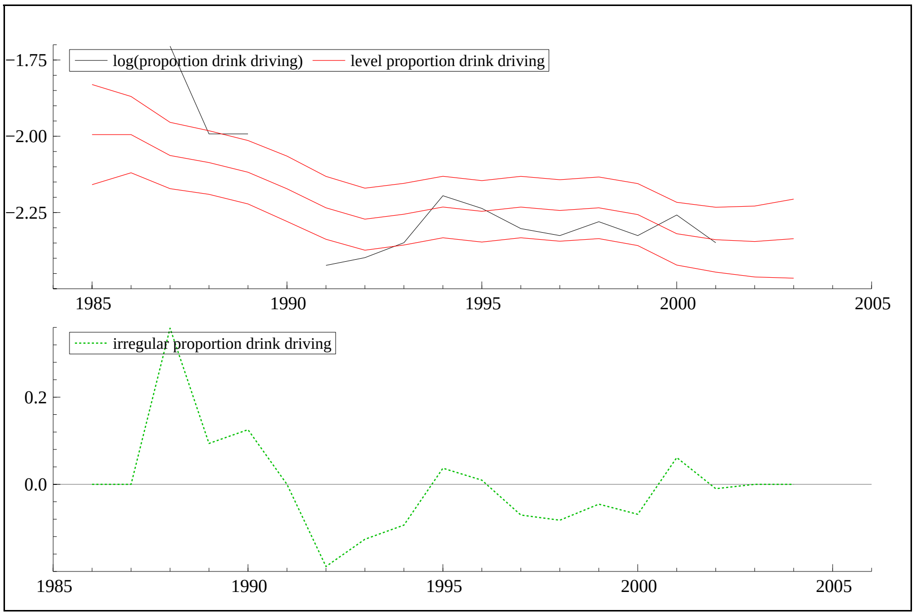

Figure 6.2. *Observed and modelled development of proportion of drink driving, and irregular component.*

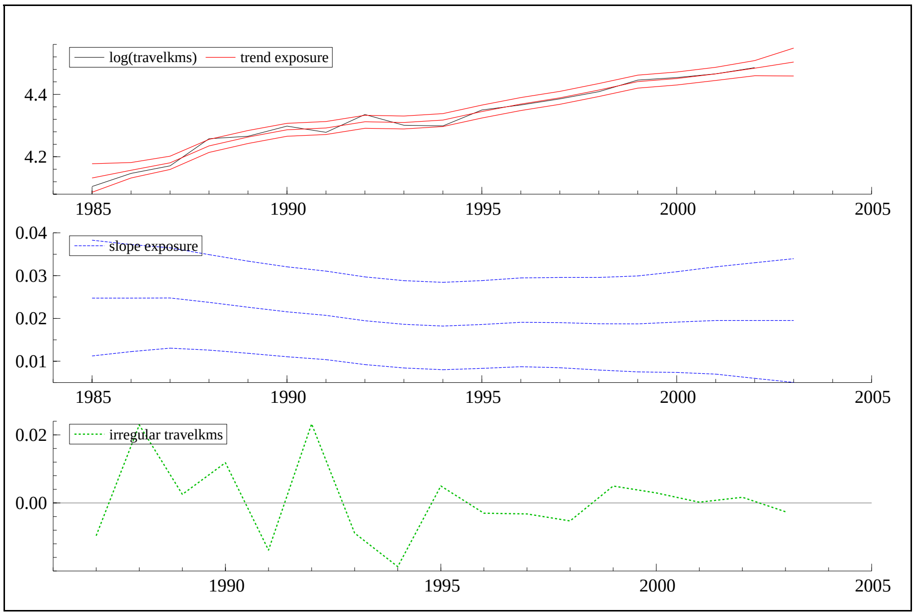

Figure 6.3. *Observed travel kilometres and modelled development of total exposure, and irregular component.*

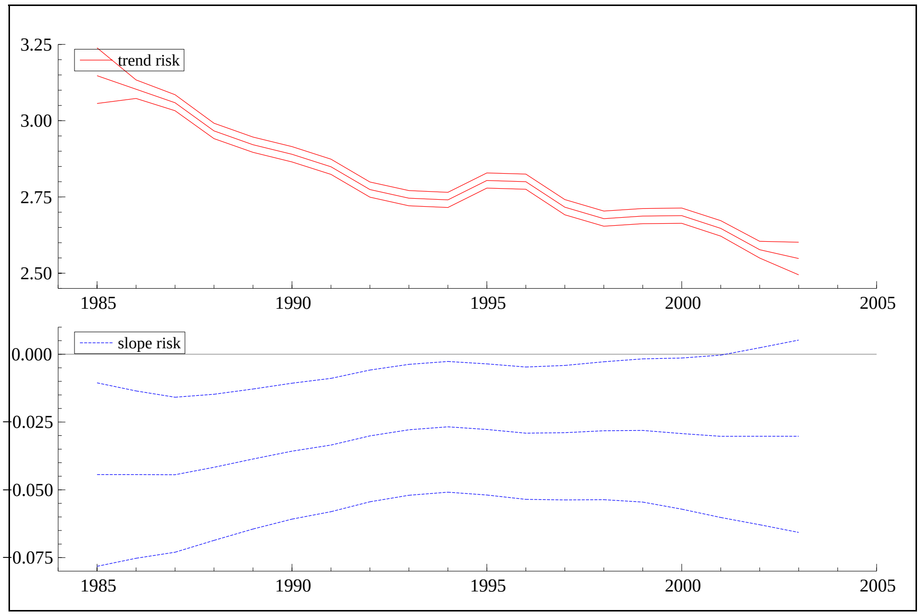

Figure 6.4. *Modelled development of total risk.*

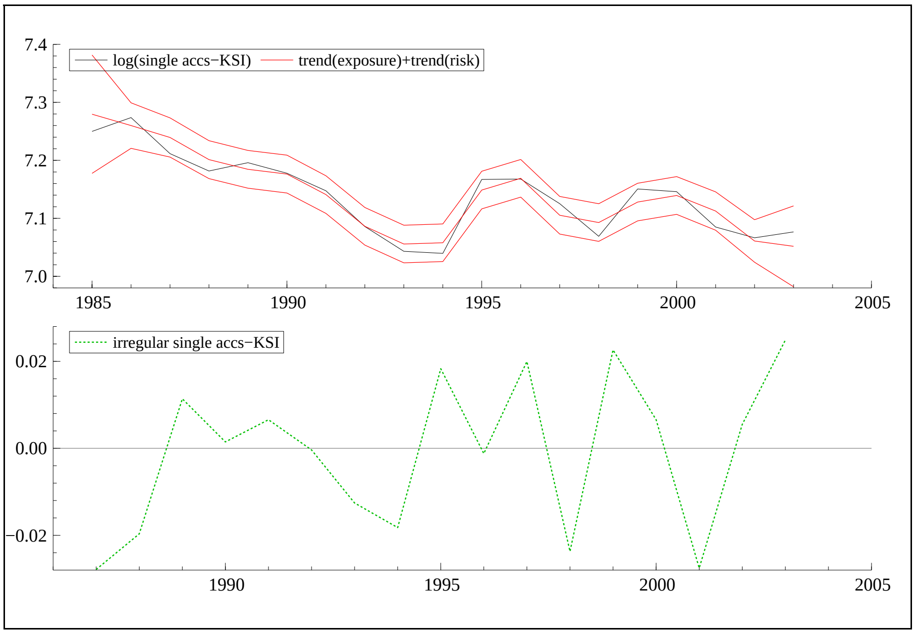

Figure 6.5. *Observed and modelled development of single accidents KSI, and irregular component.*

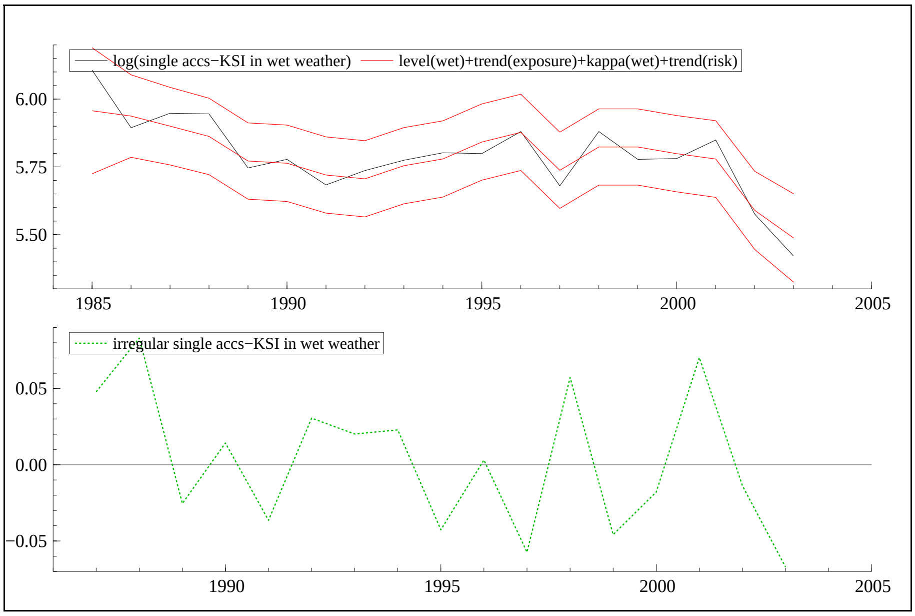

Figure 6.6. *Observed and modelled development of single accidents KSI in wet weather conditions, and irregular component.*

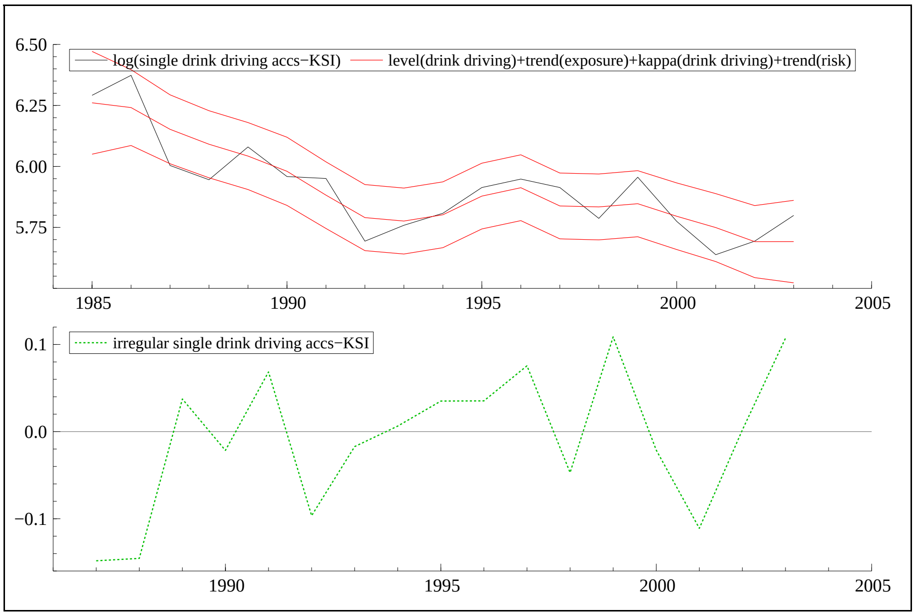

Figure 6.7. *Observed and modelled development of single drink driving accidents KSI, and irregular component.*

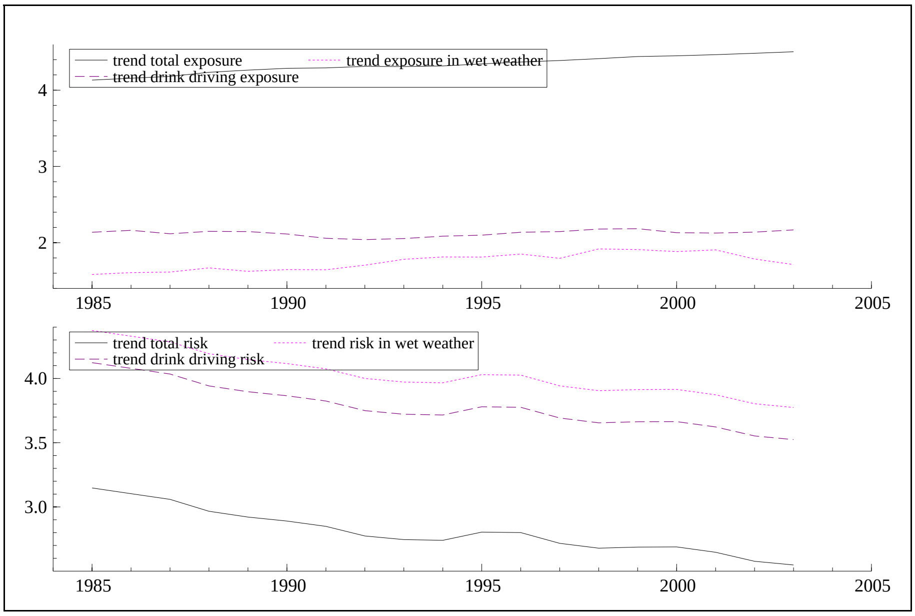

Figure 6.8. *Trends of total exposure, and of exposure in wet weather and drink driving conditions (top), and trends of total risk, and of risk in wet weather and drink driving conditions (bottom).*

The maximum likelihood estimate of log(k_t^{wet}) in equation in (11) equals 1.2261 for *t* = 1985, …, 2003, indicating an *increased* risk in wet weather conditions of $e^{1.2261}$ = 3.41. The value of log(k_t^{wet}) is very significant, since the value of the corresponding *t*-test equals 28.60. In contrast with the first method described in *Chapter 5*, the present method does result in an effect of the explanatory variable 'proportion of time with wet weather' on the risk which is in the expected direction.

The maximum likelihood estimate of log(k_t^{alc}) in equation in (15) equals 0.9758 for t = 1985, …, 2003. Since $e^{0.8552}$ = 2.65 this indicates a more than two and a half-fold increase of the risk in drink driving conditions. The value of log(k_t^{alc}) is very significant, since the value of the t-test equals 22.89. Again, this result is in line with the well-known positive relationship between drink driving and the risk of becoming involved in a road accident.

In *Figure 6.8* the modelled total exposure is shown together with the modelled exposure in wet weather, and in drink driving conditions (top). The modelled total risk together with the modelled risk in wet weather, and in drink driving conditions are displayed at the bottom of *Figure 6.8*. Clearly, the exposure in wet weather and drink driving conditions is much smaller than the total exposure, while the risk in wet weather and in drink driving conditions is considerably larger than the total risk.
The sum of the solid lines in *Figure 6.8* equals the modelled development for the total number of single accidents KSI (as shown in *Figure 6.5*). The sum of the dotted lines in *Figure 6.8* equals the modelled development for the number of single accidents KSI that occurred in wet weather conditions (as shown in *Figure 6.6*). Finally, the sum of the dashed lines in *Figure 6.8* equals the modelled development for the number of single drink driving accidents KSI (as shown in *Figure 6.7*).

The estimated effect of drink driving in the present analysis is still rather low. This is related to the fact that drink driving accidents KSI are often not reported as such. As far as severely injured drivers are concerned, the police almost always only requests a blood sample when they *suspect* the driver of drink driving; even then, in some cases the taking of a blood sample is refused by the medical hospital staff on medical grounds. Moreover, in the Netherlands, drivers who die in a road accident are not tested for alcohol, and the accident is therefore not reported as a drink driving accident.
To illustrate the impact of under-registration on the estimated effect of drink driving, a re-analysis of the data in *Table 2.1* with an assumed ratio of 1:2.5 for registered versus actual single drink driving accidents KSI results in a value of 1.8921 for log(k_t^{alc}), implying a risk increase of $e^{1.8921}$ = 6.63 in drink driving conditions.

7. Discussion and conclusions

First, results of the basic evaluation model (BEM) were shown when applied to single accidents KSI. Next, two methods were presented for incorporating explanatory variables in the BEM. In the first method the explanatory variables the ‘proportion of time with wet weather’ and the ‘proportion of drink driving’ are simply added as fixed variables to the risk component of the BEM. With the latter method, a negative relation is found between the proportion of time with wet weather and (the risk of becoming involved in) single accidents KSI, implying that (the risk of becoming involved in) single accidents KSI increase(s) as the ‘proportion of time with wet weather’ decreases. Moreover, with this method no relation is found between drink driving and risk. This indicates that the first method clearly fails to give the expected results.

In the second method, the explanatory variable ‘proportion of time with wet weather’ is not treated as fixed and known, but is assumed to be contaminated with measurement error. The latent component of the explanatory variable is then used to obtain an estimate of the proportion of exposure in wet weather conditions, and the effect of the explanatory variable on risk is ascertained by estimating how much the risk of becoming involved in single accidents KSI needs to be in- or decreased in order to optimize the prediction of the observed number of single accidents KSI *that occurred in wet weather conditions* (defined as F_t^{wet} in *Chapter 6*). The same procedure is used to ascertain the effect of the explanatory variable ‘proportion of drink driving’ on single accidents KSI.
The second method therefore requires extra information: we need to know the number of single accidents KSI that occurred in wet weather conditions, and the number of single accidents KSI involving drink driving. However, in contrast with the first method, the second method successfully uncovers the positive relations one would expect between the explanatory variables ‘proportion of time with wet weather’ and ‘proportion of drink driving’, and the risk of becoming involved in single accidents KSI.

The methodology shows how to handle time dependencies in the data (and therefore how to minimize serial correlation in the residuals). It also points out that missing observations (like the travel kilometres for 2003, and the proportions of drink driving in the years 1985, 1986, 1990, 2002, and 2003) are usually easily dealt with in state space methods.

In future research, explicit diagnostic tests will have to be added to check whether the residuals of the models do indeed satisfy independence, homoscedasticity (i.e., homogeneity of variance) and normality, since these conditions should be met (in that order of importance) for reliable significance tests of the effects of explanatory variables, and for the computation of reliable confidence intervals for the modelled developments. The analyses discussed in the present report could be improved by weighting the observed travel kilometres with their error variances, in order to decrease the heteroscedasticity (i.e., heterogeneity of variance) observed in the irregular component for this variable (see the bottom graphs in *Figures 4.1*, *5.2*, and *6.3*).

Moreover, the second method could be extended to include an evaluation of the effect of the *interaction* between wet weather and drink driving; in that case yet another time series, consisting of the annual numbers of single drink driving accidents KSI in wet weather, should be added to the model.

When estimating the effect of drink driving on road safety, the following extensions and alternatives could be considered in future research:

1. Since higher BAC percentages are associated with (much) higher risks of becoming involved in an accident, the method used in the present report (where all BAC percentages higher than 0.05 percent were considered simultaneously) could be refined by disaggregating drink driving into several classes of increasing BAC.
2. In the present report, estimates of drink driving during the whole week were derived by only considering one study (i.e., the Tilburg study). However, more studies are available on this topic. For example, a year long study was performed in Rotterdam during the whole week (Vis, 1987). The results of such studies could be used to improve or at least validate the estimates made in the present report for the proportion of drink driving in the Netherlands during the whole week.
3. An alternative to the analyses discussed in the present report would be to restrict the complete analysis to weekend nights only, for all variables involved. In this case, information would be required on the numbers of single accidents KSI and travel kilometres during weekend nights only, and on the proportion of weekend nights with wet weather conditions.
4. In order to obtain better estimates of the effect of drink driving on accidents KSI, the under-registration of drink driving accidents KSI should somehow be taken into account in future time series analyses.

References

Bijleveld, F.D. (1999). *Monitoring van verkeersveiligheid : beschrijving van een rekeninstrument voor het volgen van ontwikkelingen in de verkeersveiligheid*. R-99-20. SWOV, Leidschendam.

Bijleveld, F.D. & Commandeur, J.J.F. (2004). *The basic evaluation model*. Paper presented at the ICTSA meeting, 27-28 May 2004, INRETS, Arceuil, France. D-2006-2. SWOV, Leidschendam.

Commandeur, J.J.F. (2004). *Handling explanatory variables in the state versus the observation equation*. Paper presented at the ICTSA meeting, 27-28 May 2004, INRETS, Arceuil, France.

Doornik, J.A. (2002). *Object-oriented matrix programming using Ox 3rd ed*. Timberlake Consultants Press, London.
www.doornik.com.

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

Gould, P.G., Bijleveld, F.D. & Commandeur, J.J.F. (2004). *Forecasting road crashes: a comparison of state space models*. Paper presented at the 24th International Symposium on Forecasting, 4-7 July 2004, Sydney, Australia.

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

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

Mathijssen, M.P.M. & Houwing, S. *The prevalence and relative risk of drink and drug driving in the Netherlands – a case-control study in the Tilburg police district*. R-2005-9. SWOV, Leidschendam.

Vis, A.A. (1987). *Onderzoek naar alcohol-, geneesmiddelen- en drugsgebruik bij verkeersslachtoffers*. R-87-32. SWOV, Leidschendam.

Appendix 1 The BEM in state space notation

The basic evaluation model presented in *Chapter 3* is a special case of state space methods for the analysis of time series (Harvey, 1989; Durbin & Koopman, 2001). In matrix algebra, all state methods can very generally be written as:

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

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

for t = 1, ..., n, where (a) is called the observation or measurement equation (and ε_t is an irregular component consisting of observation errors or disturbances), and (b) is called the state equation (and η_t contains the state disturbances). The basic evaluation model for evaluating the individual developments in Dutch road safety is a bivariate local linear trend model. Specifically, let:

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

where M_t are the observed mobility figures and F_t are the observed accident figures at time points t = 1, ..., n.
Define:

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

$$R_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

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

$$Q_t = \begin{bmatrix} \sigma^2_{\xi^{(1)}} & \mathrm{cov}(\xi^{(1)},\xi^{(2)}) & 0 & 0 \\ \mathrm{cov}(\xi^{(1)},\xi^{(2)}) & \sigma^2_{\xi^{(2)}} & 0 & 0 \\ 0 & 0 & \sigma^2_{\zeta^{(1)}} & \mathrm{cov}(\zeta^{(1)},\zeta^{(2)}) \\ 0 & 0 & \mathrm{cov}(\zeta^{(1)},\zeta^{(2)}) & \sigma^2_{\zeta^{(2)}} \end{bmatrix}$$

Then writing out (a) in scalar notation yields the following two so-called observation equations:

$$\begin{aligned} y_t^{(1)} &= \mu_t^{(1)} + \varepsilon_t^{(1)} \\ y_t^{(2)} &= \mu_t^{(1)} + \mu_t^{(2)} + \varepsilon_t^{(2)} \end{aligned}, \qquad \text{(c)}$$

while working out (b) in scalar notation results in the following four so-called state equations:

$$\begin{aligned} \mu_{t+1}^{(1)} &= \mu_t^{(1)} + \nu_t^{(1)} + \xi_t^{(1)} \\ \mu_{t+1}^{(2)} &= \mu_t^{(2)} + \nu_t^{(2)} + \xi_t^{(2)} \\ \nu_{t+1}^{(1)} &= \nu_t^{(1)} + \zeta_t^{(1)} \\ \nu_{t+1}^{(2)} &= \nu_t^{(2)} + \zeta_t^{(2)} \end{aligned}. \qquad \text{(d)}$$

The two equations in (c) are identical to equations (3) and (4) in *Chapter 3*, and the four equations in (d) are identical to equations (5), (7), (6) and (8) in *Chapter 3*, respectively.

Appendix 2 The BEM with explanatory variables in state space notation, first method

To add three, say, deterministic explanatory variables to the risk component of models (a) and (b), define:

$$\alpha_t = \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \\ \mu_t^{(1)} \\ \mu_t^{(2)} \\ \nu_t^{(1)} \\ \nu_t^{(2)} \end{pmatrix}, \ \eta_t = \begin{pmatrix} \tau_{1,t} \\ \tau_{2,t} \\ \tau_{3,t} \\ \xi_t^{(1)} \\ \xi_t^{(2)} \\ \zeta_t^{(1)} \\ \zeta_t^{(2)} \end{pmatrix}, \ T_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ x_{1,t} & x_{2,t} & x_{3,t} & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R_t = I_7 ,$$

$$Z_t = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix},$$

$$H_t = \begin{bmatrix} \sigma^2_{\varepsilon^{(1)}} & \mathrm{cov}(\varepsilon^{(1)}, \varepsilon^{(2)}) \\ \mathrm{cov}(\varepsilon^{(1)}, \varepsilon^{(2)}) & \sigma^2_{\varepsilon^{(2)}} \end{bmatrix}, \text{ and}$$

$$Q_t = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix}, \text{ with}$$

$$A = \begin{bmatrix} \sigma^2_{1\tau} & 0 & 0 \\ 0 & \sigma^2_{2\tau} & 0 \\ 0 & 0 & \sigma^2_{3\tau} \end{bmatrix}, \ B = \begin{bmatrix} \sigma^2_{\xi^{(1)}} & \mathrm{cov}(\xi^{(1)}, \xi^{(2)}) \\ \mathrm{cov}(\xi^{(1)}, \xi^{(2)}) & \sigma^2_{\xi^{(2)}} \end{bmatrix}, \text{ and}$$

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

Writing out (a) in scalar notation yields the following two observation equations:

$$\begin{aligned} y_t^{(1)} &= \mu_t^{(1)} + \varepsilon_t^{(1)} \\ y_t^{(2)} &= \mu_t^{(1)} + \mu_t^{(2)} + \varepsilon_t^{(2)} \end{aligned}, \qquad \text{(e)}$$

as before, while working out (b) in scalar notation results in the following seven state equations:

$$\begin{aligned}
&\beta_{1,t+1} = \beta_{1,t} + \tau_{1,t} \\
&\beta_{2,t+1} = \beta_{2,t} + \tau_{2,t} \\
&\beta_{3,t+1} = \beta_{3,t} + \tau_{3,t} \\
&\mu_{t+1}^{(1)} = \mu_t^{(1)} + \nu_t^{(1)} + \xi_t^{(1)} \\
&\mu_{t+1}^{(2)} = \mu_t^{(2)} + \nu_t^{(2)} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t} + \beta_{3,t}x_{3,t} + \xi_t^{(2)} \\
&\nu_{t+1}^{(1)} = \nu_t^{(1)} + \zeta_t^{(1)} \\
&\nu_{t+1}^{(2)} = \nu_t^{(2)} + \zeta_t^{(2)}
\end{aligned} \quad . \qquad \text{(f)}$$

The first three state equations in (f) are needed to estimate the regression weights, while the fifth state equation in (f) represents the trend in risk including the effects of the explanatory variables on the risk (as in equation (9) in *Chapter 5*).

Appendix 3 The BEM with explanatory variables in state space notation, second method

The second method for adding explanatory variables to the basic evaluation model is a multivariate time series model where we now consider the following *six* dependent variables:

$$y_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \\ y_t^{(6)} \end{pmatrix} = \begin{pmatrix} \log(p_t^{wet}) \\ \log(p_t^{alc}) \\ \log(M_t) \\ \log(F_t) \\ \log(F_t^{wet}) \\ \log(F_t^{alc}) \end{pmatrix},$$

where

- p_t^{wet} is the observed proportion of time of the year when the weather is not dry (as measured in de Bilt) at time point *t*,
- p_t^{alc} is the observed proportion of drink driving at time point *t*,
- M_t is the observed number of kilometers travelled by cars at time point *t*,
- F_t is the observed number of single motor vehicle accidents involving people being Killed or Severely Injured (KSI) at time point *t*,
- F_t^{wet} is the observed number of single motor vehicle accidents KSI that occurred in wet circumstances (see the fourth column in *Table 2.1*) at time point *t*,
- F_t^{alc} is the observed number of single drink driving motor vehicle accidents KSI at time point *t*.

Define:

$$\alpha_t = \begin{pmatrix} \kappa_t^{(1)} \\ \kappa_t^{(2)} \\ \mu_t^{(1)} \\ \mu_t^{(2)} \\ \mu_t^{(3)} \\ \mu_t^{(4)} \\ \nu_t^{(3)} \\ \nu_t^{(4)} \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \\ \xi_t^{(3)} \\ \xi_t^{(4)} \\ \zeta_t^{(3)} \\ \zeta_t^{(4)} \end{pmatrix}, \quad T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \; Z = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix},$$

$$H_t = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \text{ with } A = \begin{bmatrix} \sigma^2_{\varepsilon(1)} & 0 \\ 0 & \sigma^2_{\varepsilon(2)} \end{bmatrix} \text{ and}$$

$$B = \begin{bmatrix} \sigma^2_{\varepsilon(3)} & \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(4)}) & \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(5)}) & \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(6)}) \\ \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(4)}) & \sigma^2_{\varepsilon(4)} & \mathrm{cov}(\varepsilon^{(4)},\varepsilon^{(5)}) & \mathrm{cov}(\varepsilon^{(4)},\varepsilon^{(6)}) \\ \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(5)}) & \mathrm{cov}(\varepsilon^{(4)},\varepsilon^{(5)}) & \sigma^2_{\varepsilon(5)} & \mathrm{cov}(\varepsilon^{(5)},\varepsilon^{(6)}) \\ \mathrm{cov}(\varepsilon^{(3)},\varepsilon^{(6)}) & \mathrm{cov}(\varepsilon^{(4)},\varepsilon^{(6)}) & \mathrm{cov}(\varepsilon^{(5)},\varepsilon^{(6)}) & \sigma^2_{\varepsilon(6)} \end{bmatrix}$$

, and

$$Q_t = \begin{bmatrix} C & 0 & 0 & 0 \\ 0 & D & 0 & 0 \\ 0 & 0 & E & 0 \\ 0 & 0 & 0 & F \end{bmatrix} \text{ with } C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \; D = \begin{bmatrix} \sigma^2_{\xi(1)} & 0 \\ 0 & \sigma^2_{\xi(2)} \end{bmatrix},$$

$$E = \begin{bmatrix} \sigma^2_{\xi(3)} & \mathrm{cov}(\xi^{(3)},\xi^{(4)}) \\ \mathrm{cov}(\xi^{(3)},\xi^{(4)}) & \sigma^2_{\xi(4)} \end{bmatrix}, \text{ and}$$

$$D = \begin{bmatrix} \sigma^2_{\zeta(3)} & \mathrm{cov}(\zeta^{(3)},\zeta^{(4)}) \\ \mathrm{cov}(\zeta^{(3)},\zeta^{(4)}) & \sigma^2_{\zeta(4)} \end{bmatrix}$$

The matrix equation (a) now consists of the following six measurement equations:

$$
\begin{aligned}
y_t^{(1)} &= \mu_t^{(1)} + \varepsilon_t^{(1)} \\
y_t^{(2)} &= \mu_t^{(2)} + \varepsilon_t^{(2)} \\
y_t^{(3)} &= \mu_t^{(3)} + \varepsilon_t^{(3)} \\
y_t^{(4)} &= \mu_t^{(3)} + \mu_t^{(4)} + \varepsilon_t^{(4)} \\
y_t^{(5)} &= \mu_t^{(1)} + \mu_t^{(3)} + \mu_t^{(4)} + \kappa_t^{(1)} + \varepsilon_t^{(5)} \\
y_t^{(6)} &= \mu_t^{(2)} + \mu_t^{(3)} + \mu_t^{(4)} + \kappa_t^{(2)} + \varepsilon_t^{(6)}
\end{aligned}
, \qquad \text{(g)}
$$

where

- $\mu_t^{(1)}$ is the unobserved trend for the proportion of time with wet weather,
- $\mu_t^{(2)}$ is the unobserved trend for the proportion of drink driving
- $\mu_t^{(3)}$ is the unobserved trend for the total exposure,
- $\mu_t^{(4)}$ is the unobserved trend for the total risk,
- $\kappa_t^{(1)}$ is the total risk changing factor due to wet weather conditions,
- $\kappa_t^{(2)}$ is the total risk changing factor due to drink driving, and
- $\varepsilon_t^{(i)}$ are white noise terms for *i* = 1, …, 6.

Matrix equation (b) now consists of the following eight state equations, which represent two local level models (for the proportion of time with wet weather, and for the proportion of drink driving) and two local linear trend models (for exposure and risk):

$$
\begin{aligned}
\kappa_{t+1}^{(1)} &= \kappa_t^{(1)} \\
\kappa_{t+1}^{(2)} &= \kappa_t^{(2)} \\
\mu_{t+1}^{(1)} &= \mu_t^{(1)} + \xi_t^{(1)} \\
\mu_{t+1}^{(2)} &= \mu_t^{(2)} + \xi_t^{(2)} \\
\mu_{t+1}^{(3)} &= \mu_t^{(3)} + \nu_t^{(3)} + \xi_t^{(3)} \\
\mu_{t+1}^{(4)} &= \mu_t^{(4)} + \nu_t^{(4)} + \xi_t^{(4)} \\
\nu_{t+1}^{(3)} &= \nu_t^{(3)} + \zeta_t^{(3)} \\
\nu_{t+1}^{(4)} &= \nu_t^{(4)} + \zeta_t^{(4)}
\end{aligned}
. \qquad \text{(h)}
$$

The six observation equations in (g) are identical to equations (10), (14), (3), (4), (11), and (15) discussed in *Chapters 3* and *6*, respectively. The eight state equations in (h) are identical to equations (12), (16), (13), (17), (5), (7), (6), and (8) discussed in the same chapters, respectively.