

Model-based measurement of latent risk in time series

Misurazione del rischio latente mediante modelli di serie temporali

F. Bijleveld¹, J. Commandeur¹, P. Gould², S. J. Koopman³

¹ SWOV Institute for Road Safety Research, Netherlands

² Department of Econometrics and Business Statistics, Monash University, Australia

³ Department of Econometrics, Vrije Universiteit Amsterdam, Netherlands

e-mail: s.j.koopman@feweb.vu.nl

Riassunto: La valutazione del rischio assume un ruolo centrale nelle decisioni di imprese, governi e altre istituzioni: il rischio di incidente stradale concerne i governi locali, quello di default della controparte concerne gli istituti di credito, mentre quello di infezione ha implicazioni di carattere attuariale ed epidemiologico. Tuttavia, il rischio non è direttamente osservabile e generalmente varia nel tempo. Il lavoro si occupa della misurazione del rischio nel contesto di un modello multivariato di serie temporali che incorpora il rischio come processo latente. Il modello per il rischio estende gli approcci esistenti per la modellizzazione congiunta de (i) l'esposizione ad un evento, (ii) il rischio di ricorrenza dell'evento e (iii) la gravità dell'evento. In primo luogo, vengono discussi gli approcci esistenti per l'analisi del rischio che sono stati applicati alla sicurezza stradale e a problemi attuariali e epidemiologici. In secondo luogo, presentiamo un modello generale per l'analisi del rischio e discuteremo il suo trattamento statistico utilizzando la metodologia state space. Quindi, applicheremo la metodologia a serie temporali concernenti le polizze assicurative, gli acquisti mediante carte di credito e la sicurezza stradale. Si conclude che la metodologia proposta si rivela assai efficace per la valutazione del rischio.

Keywords: Actuarial statistics; Dynamic factor analysis; Kalman filter; Log-odd ratios; Maximum likelihood; Road casualties; State space model; Unobserved components.

1. Introduction

In the literature the term “risk” may take many meanings. In financial time series literature, realized and implied volatility models treat risk as the standard deviation of returns. Governments and financial institutions are concerned with other types of risk, where particular events are studied. “Insurance risk” is modelled by actuaries who wish to measure the risk of a claim. Epidemiologists study “medical risk”, which is the risk of infection or injury. For banks, “default risk” is of great importance for regulatory and internal capital management purposes. Similarly accident researchers and credit portfolio managers are also concerned with the risk of certain events occurring, and also their likely severity. These “event risk” analyses have common elements: there is (i) exposure to risk, (ii) the risk (chance, probability) of the event occurring and (iii) the severity of the event.

The time series modelling of “event risk” offers new insights into data and can confirm or reject the validity of constant risk assumptions. There is growing pressure to develop such models in a range of fields. The Basel Accord (BIS, 2004) requires banks to be able to understand the present and also forecast the future value-at-risk of their credit

portfolios. Greater regulatory capital burdens will be placed on banks who cannot demonstrate they have robust models for default risk and associated losses. Road safety researchers now have considerable pressure from governments to be able to evaluate past safety measures and forecast future accidents and injuries (WHO, 2004). The increased availability of data and continuing improvements in computer power have also opened up a range of new models which can be applied to time series data.

There is substantial evidence that simple deterministic models fail to adequately explain the dynamics of actuarial and epidemiological systems. Recently a number of articles have examined stochastically time-varying structures to model risk in epidemiological applications. For example, Dominici et al. (2004) find evidence of time varying risk factors within a generalised additive model framework used to determine the interaction between mortality rates and air pollution concentrations. The data is from 1987–1994, indicating time variation over relatively short time frames. In this approach, the natural log of fatality counts is modelled as a smooth semiparametric function of time and weather variables and a linear function of pollution levels. Finkenstadt and Grenfell (2000) find evidence of seasonal time variation in the transmission parameter for a susceptible-exposed-infected-recovered (SEIR) model for measles epidemics.

In actuarial research, there is a surprising lack of time series models for the rate and severity of insurance claims. Among the few articles is de Jong and Boyle (1983), in which Bayesian methods are applied to a state space model which produces stochastically time-varying mortality rates. Exposure is measured by the size of the population and is effectively treated as a covariate. While multiple population cohorts are examined, the approach does not define the dynamics of exposure and severity. Risk is defined as the time-varying, locally polynomial probability for a binomial distribution with the number of trials given by the population size. Harvey and Fernandes (1989) also develop a model for insurance claims using the stochastic unobserved components (UC) framework (Harvey, 1989), where both the size of claims and the number of claims are modelled. The focus of the Harvey and Fernandes (1989) paper is on the modelling of Poisson and normal variables that are combined to form a loss function, which is the dollar value of all claims. Automobile insurance claims for multiple cohorts are analysed by Ledolter et al. (1990), who use stochastic UC models to test for common factors (“shrinkage”) across cohorts. These models provide better fit to the data than static or deterministic models.

In bank risk management there have also been a number of articles examining the use of time varying parameters to model the risk of counterparty default. Allen and Saunders (2003) highlight the need for dynamic approaches to modelling company default. Structured financial risk models have been used successfully in this area. Multivariate approaches are often used to model observations from different cohorts, rather than separate “dimensions” of risk such as exposure to default and the number of defaults. A time-varying logistic model estimated via Kalman filtering is introduced by Fahrmeir and Wagenpfeil (1996) for duration models which assess the probability of subjects entering or leaving a state of unemployment. In this case the model is for individuals, and time is the only measure of exposure, so a logistic transform is used to produce time varying probabilities. The results suggest there is a need for time variation in model parameters to accurately model unemployment dynamics.

In road safety research the *demande routière, des accidents et leur gravité* (DRAG) framework of Gaudry (1984) and Gaudry and Lassarre (2000) has been widely applied. The DRAG framework models the dimensions of exposure, risk and severity but does not

use a multivariate UC structure. As in most “event risk” models, exposure is a covariate used to explain accident counts. The DRAG model differs from other structures in that it is usually estimated using ARMA regression models with many social and economic explanatory variables. Box-Cox transformations are applied to all the data, meaning that an extra parameter needs to be estimated for each of the dependent variables and the covariates, which often leads to models with very large parameter sets. The DRAG model’s reliance on economic and social variables, such as unemployment and alcohol consumption (which are difficult to forecast in their own right), means that it is of limited value in a forecasting context.

Despite the many common features of the data and methods applied, to date there is no single unified framework for modelling “event risk”. In this paper we introduce a general multivariate model for “event risk” analysis that can consider exposure, risk and severity simultaneously. The latent risk time series (LRT) model can be applied to a range of problems involving “event risk” and is not specifically limited to either actuarial or epidemiological applications. The standard approaches to risk treat exposure as a fixed and known variable and few models allow for the modelling of severity at all. The LRT model is general and allows for the stochastic evolution of exposure, risk and severity over time. It extends previous work by treating exposure and severity as an integral part of the risk problem. The logarithmic transformation is used to model multiplicative, time-varying relationships between exposure, risk and severity. In existing approaches some or all of these variables (particularly exposure) are treated as known, when in reality they are often measured under error and subject to stochastic variation. The LRT model has a multivariate structure and therefore correlations between latent processes and errors can be estimated. Since the model belongs to the class of UC time series models, the multivariate decomposition can include latent factors for trend, seasonal and cyclical dynamics. This general framework also allows for easy forecasting of future losses, events and exposures together with prediction confidence bounds, which are of particular interest to risk managers. The relative simplicity of the model structure means that complex estimation techniques are usually not required. Bayesian and classical estimation methods can be easily applied and usually rely on Kalman filter methods. The multivariate framework is general enough to allow for a study of multiple cohorts (similar to Ledolter et al. (1990)) with a view to enabling “shrinkage” of the number of states required.

The statistical framework, including state space forms and estimation methods are presented in Section 2. The exposure-risk motor vehicle insurance model is the first example of a LRT analysis and is discussed in Section 3. The exposure-risk-severity model for credit card use is treated in Section 4. The multiple exposure-single risk model for bicycle and moped road traffic accidents is presented in Section 5. The LRT analysis includes parameter estimation, measurement of risk and severity based on signal extraction methods and forecasts. Section 6 concludes.

2. The statistical framework

The latent risk time series (LRT) model includes latent factors for exposure, risk and severity which are all associated with the observed variables:

- x_{it} , exposure at time t for group i with $i = 1, \dots, k_x$

- y_{it} , outcome at time t for group i with $i = 1, \dots, k_y$
- z_{it} , loss at time t for group i with $i = 1, \dots, k_z$,

for $t = 1, \dots, n$ where n is the number of observations. There are k_x , k_y and k_z groups for the exposure, outcome and loss variables, respectively. There is no need to set $k_x = k_y = k_z$ because multiple outcomes for only one exposure variable (such as multiple types of vehicle accident per registered vehicle) can exist so that $k_x < k_y$.

The exposure variable may take many forms. For example, exposure to traffic crash risk may be measured in terms of vehicle registrations or by distance travelled. For a bank, the exposure variable may be either the number of loans or the dollar value of loans. The exposure is to the risk of a particular event or outcome occurring. The outcome variable y_{it} is typically the number of times a certain event occurs within group i . In insurance applications this may be the number of claims. For epidemiologists the outcome variable may be the number of successful treatments. Indeed, the outcome variable does not have to be undesirable.

In some cases a loss variable is also required to fully describe the risk associated with a particular outcome. The loss variable measures the severity (or consequences) of the outcomes. For an insurance company the loss variable may be the dollar value of claims. The severity of a loss is also important for bank risk managers who wish to know how much money is likely to be lost in the event of default.

2.1. The LRT model

As in the modelling of disease incidence by Morton and Finkenstadt (2005), unobserved components or latent processes are required to properly define a risk system for the observed variables exposure, outcome and loss. The LRT model relates these observed variables within a multivariate system of equations:

$$\begin{aligned} x_{it} &= E_{it} \times U_{it}^{(x)}, \\ y_{it} &= E_{it} \times R_{it} \times U_{it}^{(y)}, \\ z_{it} &= E_{it} \times R_{it} \times S_{it} \times U_{it}^{(z)}, \end{aligned}$$

where E_{it} , R_{it} , and S_{it} are the latent variables exposure, risk and severity for group i at time t , respectively. The error terms $U_{it}^{(x)}$, $U_{it}^{(y)}$ and $U_{it}^{(z)}$ reflect that observed variables are measured under uncertainty. For example, road-use surveys may not be accurately reported or there may simply be unexplainable short-term deviations from the trend.

After taking logs, the multiplicative LRT equations can be expressed in additive form:

$$\begin{aligned} \log x_{it} &= \mu_{it}^{(E)} + \varepsilon_{it}^{(x)}, \\ \log y_{it} &= \mu_{it}^{(E)} + \mu_{it}^{(R)} + \varepsilon_{it}^{(y)}, \\ \log z_{it} &= \mu_{it}^{(E)} + \mu_{it}^{(R)} + \mu_{it}^{(S)} + \varepsilon_{it}^{(z)}, \end{aligned} \tag{1}$$

where $\mu_{it}^{(E)} = \log E_{it}$ is log-exposure, $\mu_{it}^{(R)} = \log R_{it}$ is log-risk and $\mu_{it}^{(S)} = \log S_{it}$ is log-severity for group i and time t . The additive noise terms $\varepsilon_{it}^{(x)} = \log U_{it}^{(x)}$, $\varepsilon_{it}^{(y)} = \log U_{it}^{(y)}$ and $\varepsilon_{it}^{(z)} = \log U_{it}^{(z)}$ are identical independently distributed (i.i.d) with mean zero but their distributions do not need to be determined necessarily. The latent factors of exposure, risk and severity have time subscripts and the form of time-variation are established next.

Unobserved components (UC) models provide a flexible framework for allowing both deterministic and stochastic variation in the latent factors. A possible UC model to describe the evolution of the latent factors is the multivariate local linear trend (LLT) specification, as used by Harvey and Fernandes (1989) in their study of insurance data. The LLT model for the log-exposure factor is given by $\mu_t^{(E)}$ are given by

$$\begin{aligned}\mu_t^{(E)} &= \mu_{t-1}^{(E)} + \delta_{t-1}^{(E)} + \eta_t^{(E)}, & \eta_t^{(E)} &\sim \text{i.i.d}(0, \sigma_{\eta^{(E)}}^2), \\ \delta_t^{(E)} &= \delta_{t-1}^{(E)} + \zeta_t^{(E)}, & \zeta_t^{(E)} &\sim \text{i.i.d}(0, \sigma_{\zeta^{(E)}}^2),\end{aligned}\quad (2)$$

for $t = 1, \dots, n$ and with the subscript i suppressed for convenience of notation. The trend component $\mu_t^{(E)}$ is a random walk process with stochastic drift $\delta_t^{(E)}$ and trend innovation $\eta_t^{(E)}$ where the latter is i.i.d with mean zero and variance $\sigma_{\eta^{(E)}}^2$. The drift term is modelled as a random walk process with i.i.d innovation $\zeta_t^{(E)}$ that has mean zero and variance $\sigma_{\zeta^{(E)}}^2$. The trend and slope innovations are mutually uncorrelated at all time points. Some special cases of the LLT model are the linear trend function ($\sigma_{\eta^{(E)}}^2 = \sigma_{\zeta^{(E)}}^2 = 0$), the random walk ($\delta_t^{(E)} = 0$ for all t), the random walk with fixed drift ($\sigma_{\zeta^{(E)}}^2 = 0$) and the integrated random walk or smooth trend ($\sigma_{\eta^{(E)}}^2 = 0$). An important assumption is that all additive errors in the equations are serially uncorrelated. The LLT model can also be considered for log-risk $\mu_t^{(R)}$ and log-severity $\mu_t^{(S)}$.

Seasonal and cyclical components can also be included in the observation equations of the LRT model (1). The seasonal component can be modelled by a series of trigonometric functions while the cyclical component can be specified as a autoregressive process with complex roots in its polynomial, see Harvey (1989, §2.3.4) for more details on the dynamic specification of unobserved components. The seasonal and cyclical processes vary stochastically over time and are usually assumed to be driven by innovations from a normal density. Furthermore, regression effects (based on covariates or explanatory variables) can be included in the model. Such deterministic but unknown effects can be added to the model when significant deviations from the exposure-risk-severity relationships can be captured by specific covariates.

For a bivariate exposure-risk model, seasonal or cyclical variations in road use may be viewed as persistent periodic (quarter-to-quarter, month-to-month) changes in exposure. For example, seasonal components for exposure and risk may be added to the measurement equations for exposure and outcome, that is

$$\begin{aligned}\log x_t &= \mu_t^{(E)} + \gamma_t^{(E)} + \varepsilon_t^{(x)}, \\ \log y_t &= \mu_t^{(E)} + \gamma_t^{(E)} + \mu_t^{(R)} + \gamma_t^{(R)} + \varepsilon_t^{(y)},\end{aligned}\quad (3)$$

where $\gamma_t^{(E)}$ and $\gamma_t^{(R)}$ are the seasonal components for exposure and risk, respectively. In this case, the seasonal factors are an integral part of exposure and risk effects in this formulation. Alternatively, the seasonal components can be exclusively attributed to particular observation equations. A particular seasonal component is then added exclusively to an observation equation while the other observation equations are not affected by this component. The alternative formulation with seasonal components is then given by

$$\begin{aligned}\log x_t &= \mu_t^{(E)} + \gamma_t^{(x)} + \varepsilon_t^{(x)}, \\ \log y_t &= \mu_t^{(E)} + \mu_t^{(R)} + \gamma_t^{(y)} + \varepsilon_t^{(y)}.\end{aligned}\quad (4)$$

where $\gamma_t^{(x)}$ and $\gamma_t^{(y)}$ represent independent seasonal time series processes. The seasonal variation in $\log y_t$ is captured by $\gamma_t^{(y)}$ rather than $\gamma_t^{(E)} + \gamma_t^{(R)}$. As a consequence, only $\mu_{it}^{(E)}$ represents log-exposure in the observation equation for $\log y_t$. The same discussion of alternative LRT model specifications applies to cyclical components and regression effects.

2.2. State space analysis

The unobserved components equations of the LRT model can be written in state space form. The general state space form for a vector of observed variables is given by

$$\mathbf{y}_t = \mathbf{Z}_t \alpha_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim \text{i.i.d}(0, \mathbf{H}_t), \quad (5)$$

$$\alpha_{t+1} = \mathbf{T}_t \alpha_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim \text{i.i.d}(0, \mathbf{Q}_t), \quad (6)$$

where α_t is the state vector, $\mathbf{e}_t$ is the disturbance vector of the observation equation (5) and $\mathbf{u}_t$ is the disturbance vector of the state equation (6). The disturbance vectors are mutually and serially not correlated at all time periods. The system matrices of the state space form are $\mathbf{Z}_t$, transition matrix $\mathbf{T}_t$ and the variance matrices $\mathbf{H}_t$ and $\mathbf{Q}_t$. The system matrices can be time-varying but are usually time invariant and sparse matrices. Some elements of the system matrices are unknown and are treated as parameters that can be estimated by maximum likelihood.

2.3. Measurement and estimation

The state space framework contains two unknown entities, the state vector α_t and the parameter vector ψ that contains all unknown elements in the system matrices. The state vector can be predicted conditional on observations. Filtering refers to the estimation of α_t conditional on $\mathbf{y}_1, \dots, \mathbf{y}_t$, that is all observations upto and including $\mathbf{y}_t$. Smoothing is similar but the estimation is conditional on all observations $\mathbf{y}_1, \dots, \mathbf{y}_n$. Filtering and smoothing methods also compute standard errors of the estimates. The Kalman filter and related methods carry out the necessary computations for the linear state space model (5) – (6). In case the disturbances are normally distributed, we obtain minimum mean squared estimators. When normality must be rejected, they are minimum mean squared linear estimators.

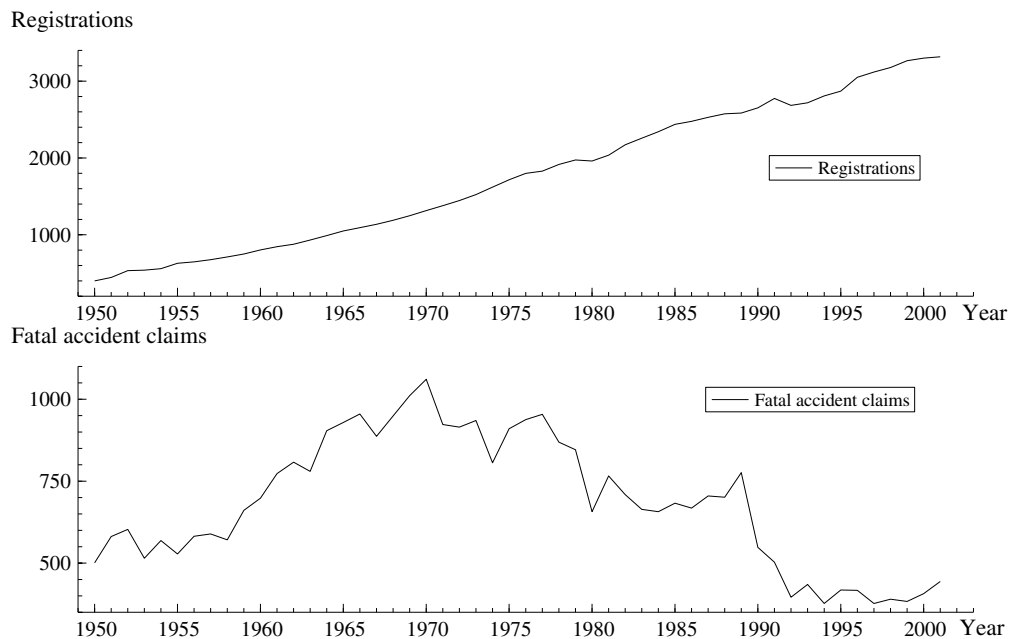
The Kalman filter carries out the prediction error decomposition for a given state space model (5) – (6) for a particular value of ψ . This implies that the likelihood function is evaluated by the Kalman filter for a given ψ . Maximum likelihood estimation of ψ is therefore a standard exercise of numerically maximising the likelihood function with respect to ψ and for which the Kalman filter is used to evaluate the likelihood values for different ψ 's in a computationally efficient way, see Harvey (1989). To ensure positive semi-definite variance matrices, the original unrestricted variance matrix Σ_η is transformed by $\Sigma_\eta = M'M$ where M is a particular symmetric matrix. Estimation concentrates on M . Monte Carlo methods are employed to produce confidence intervals for the parameter estimates. For this purpose, matrices M are simulated from the multivariate normal distribution with mean $\text{vech}(\widehat{M})$ and variance matrix V where $\widehat{M}$ is the maximum likelihood estimate of M and V equals the negative inverse of the numerical second derivative of the likelihood function with respect to $\text{vech}(M)$ and evaluated at $M = \widehat{M}$. Elements of M that tend to zero during the estimation process are fixed at zero

when some benchmark is reached.

3. A two-dimensional insurance LRT model

A two-dimensional LRT model is considered for the measurement of exposure and risk in relation to motor vehicle fatality insurance claims. An annual dataset from Victoria, Australia is analysed for this purpose. Victoria vehicle registrations represent the number of policies (exposure) for the compulsory third-party insurance body, the Transport Accident Commission (TAC). The fatality series is the outcome variable and represents the numbers of a certain type of claims on these policies for different years. Although the TAC was created in 1986, the data used to fit the model is from 1950–2001. The earlier observations are valid for parameter estimation as they represent the same portfolio, which is the state’s entire vehicle population. The LRT model disentangles risk and exposure effects from observations of motor vehicle fatalities and vehicle registrations. This type of data has previously been analysed using the Oppe model (Oppe, 1989), which assumes exposure follows a logistic-S curve and log-risk evolves deterministically and typically with a downwards drift.

Figure 1: *Victorian registrations and vehicle crash fatalities 1950–2001.*



The observed series are presented in Figure 1. The exposure series, registrations, displays an upwards trend and is generally smooth. This is typical of the evolution of vehicle registrations in developed economies. The fatal claims series has a “hump” shape, with a peak during the start of the 1970s, which is also typical of developed economies. Because registrations have increased monotonically over the past 50 years, the reduction in fatal claims has been caused by a decrease in risk. Risk reductions have been driven by gradual improvements in vehicle and road design together with increased

public awareness. Demographic factors may also be important as a new generation of road users ("baby boomers") began to start driving. Public horror at a road toll of 1034 in 1970 led to newspaper declarations of "war on 1034", indicative of these changing attitudes towards road safety. The effects on attitude have proved to be long-term and not only instantaneous. Other important relevant events in the sample are the introduction of seatbelt laws in 1980 and the increased enforcement of mass media advertising campaigns on road safety and public safety consciousness in the early 1990s.

Because the data is not separated into groups or cohorts, there are only two observed series: policies x_t and claims y_t (the group index i is suppressed from the notation). In many insurance examples, separate cohorts for age or gender are used to model risk and exposure for different policy holders. If cohorts were used to separate male and female policy holders, we would require 4 observed series (2 for exposure and 2 for fatalities) to produce estimates of 2 exposure and 2 risk processes. Furthermore, a three-dimensional LRT can be considered to model the severity of the claims. In this case a third observed variable (the dollar value of payouts on claims) is needed to explain the expected cost per claim (severity).

In this section we focus on the two-dimensional and single cohort structure described by

$$\begin{aligned} \text{policies}_t &= x_t = E_t \times U_t^{(x)} \\ \text{claims}_t &= y_t = E_t \times R_t \times U_t^{(y)}, \end{aligned} \quad (7)$$

where E_t is exposure, R_t is risk and $U_t^{(x)}$ and $U_t^{(y)}$ are observation noise terms for policies and claims, respectively. The observations are subject to the log-transformation and the unobserved components $\mu_t^{(E)} = \log E_t$ and $\mu_t^{(R)} = \log R_t$ are flexibly modelled by the LLT specification (2). Covariates in the form of dummy variables for special events are introduced in the level ($\mu_t^{(E)}$ and $\mu_t^{(R)}$) and drift ($\delta_t^{(E)}$ and $\delta_t^{(R)}$) equations. The following events are considered: (i) in 1970, publicity started to have a safer attitude in traffic ("war on 1034"); (ii) in 1971, introduction of seatbelt laws; (iii) in 1980, change in data collection on vehicle registrations; (iv) in 1990, enforcement of advertising initiatives; (v) in 1992, another change in data collection on vehicle registrations. The changes in data collection should only affect exposure E_t while the other events should have an effect on risk R_t . The change of attitude in traffic is a long-term effect and therefore captured by a change in the drift term of risk. The effects of seatbelt laws and intensified road safety advertisements are taken as immediate step changes in the level of risk. Other interventions can also be considered but they have proved to be less important in the analysis.

Parameter estimates and simulated asymmetric 95% confidence intervals are presented in Table 1. The estimate of the variance of the observation noise for the registrations equation (policies) is closer to zero than for the claims equation. The data collection of registrations may be done more accurately or the time series of claims is just subject to more observation noise, see Figure 1. An interesting result from this analysis is that the level processes for exposure and risk are perfect negatively correlated and that the drift processes are also perfect negatively correlated. In other words, the latent variables exposure and risk are driven by two univariate noise sequences for level η_t^c and drift ζ_t^c . It

follows that

$$\begin{aligned}\mu_t^{(E)} &= \mu_{t-1}^{(E)} + \delta_t^c + \eta_t^c, & \delta_t^c &= \delta_{t-1}^c + \zeta_t^c, \\ \mu_t^{(R)} &= \mu_{t-1}^{(R)} + a\delta_t^c + b\eta_t^c,\end{aligned}$$

for coefficients $a < 0$ and $b < 0$ with common drift term δ_t^c and common noise terms η_t^c and ζ_t^c for level and drift, respectively. The perfect negative correlations mean that both exposure and risk components are subject to the same stochastic shocks that determine their time-varying behaviour. This finding is in agreement with most road crash research which finds a strong negative relationship between risk and exposure. The perfect correlation of shocks imply that the components can be interpreted as common factors. Nevertheless, the estimated components are distinct from each other since they are also subject to a number of interventions that are captured by dummy variables associated with special events. The interventions allow the components to have separate shocks or breaks despite the fact that the random shocks are perfectly correlated. The regression estimates of these intervention coefficients are summarised in Table 2. The estimated intervention for the anticipated break in the level of exposure due to a change in the data collection of policies (registrations) is clearly significant for 1992 but less significant for 1980. The level interventions for risk in 1971 (seatbelt laws) and 1990 (advertising initiatives) are very significant. Finally, the drift intervention for risk in 1970 (change in public safety consciousness initiated by government and media) is also very significant. All estimated interventions have negative values.

Figure 2 presents the estimated level and drift components of exposure and risk (in logs). The estimated components are subject to both random shock and interventions. The salient features of the analysis is the increasing exposure with a significant drift term throughout the sample, and the decreasing risk with a significant negative drift term that is mainly caused by the publicity intervention. Risk displays relatively more stochastic variation in both the estimated level and drift terms. The level also has a relatively large variance estimate of 0.00130. It is interesting to detect in Figure 2 that, apart from the intervention shocks, level and drift components of risk are perfectly and negatively correlated with level and slope components of exposure, respectively.

After a period of statistical insignificance, the slope of risk becomes significantly negative after interventions in the early 1970s. The advertising and enforcement initiative of the early 1990s were also highly effective in reducing risk. Major variations in risk are driven by interventions together with some relatively smooth stochastic variation. Since the early 1990s, log-risk has settled to a relatively steady level, with some evidence of a recent increase. Confidence intervals for risk are wider than for exposure. It indicates that most of the variation in claims is risk-driven. The intervention breaks affect periods where there is a breakdown in the perfect correlation. A clear example is the estimated drift terms that are perfect negatively correlated. The seat belt law shift of 1971 in the drift equation of risk means that the estimated slope terms appear to evolve quite differently. The interventions provide therefore some flexibility in the potentially restrictive assumptions of common levels and drifts.

Figure 3 shows that the one-step ahead prediction residual series for registrations and claims are reasonably well-behaved. Normality assumptions may be considered. The residual series of claims appear to have a significant cyclical autocorrelation pattern. Extended versions of the LRT using cycles or autoregressive components can be examined to eliminate this autocorrelation. The cyclical pattern may be related to economic

		exposure	risk
Trend μ_t	exposure	0.00031 (0.00002, 0.00111)	<i>-1.00000</i> *
	$\hat{\Sigma}_\eta$ risk	-0.00064 (-0.00204, -0.00002)	0.00130 (0.00008, 0.00415)
Slope δ_t	exposure	0.00004 (0.00001, 0.00009)	<i>-1.00000</i> *
	$\hat{\Sigma}_\zeta$ risk	-0.00007 (-0.00018, -0.00001)	0.00013 (0.00001, 0.00039)
		policies	claims
Noise ε_t	policies	0.00016 (0, 0.00056)	<i>0</i> *
	$\hat{\Sigma}_\varepsilon$ claims	0 *	0.004206 (0.00252, 0.00629)

Table 1: Victorian crash data: ML estimates of variance matrices with asymmetric 95% confidence intervals in parentheses. Correlations are given in italics on the upper triangular elements of the variance matrices.

equation	year	event	estimate	t-stat
$\delta_t^{(R)}$	1970	war on 1034	-0.0785	-3.79
$\mu_t^{(R)}$	1971	seatbelt law	-0.1084	-4.66
$\mu_t^{(E)}$	1980	change in data collection	-0.0864	-2.04
$\mu_t^{(R)}$	1990	advertising initiatives	-0.3757	-6.74
$\mu_t^{(E)}$	1992	change in data collection	-0.0662	-8.63

Table 2: Victorian crash data: Intervention estimates with t-statistics

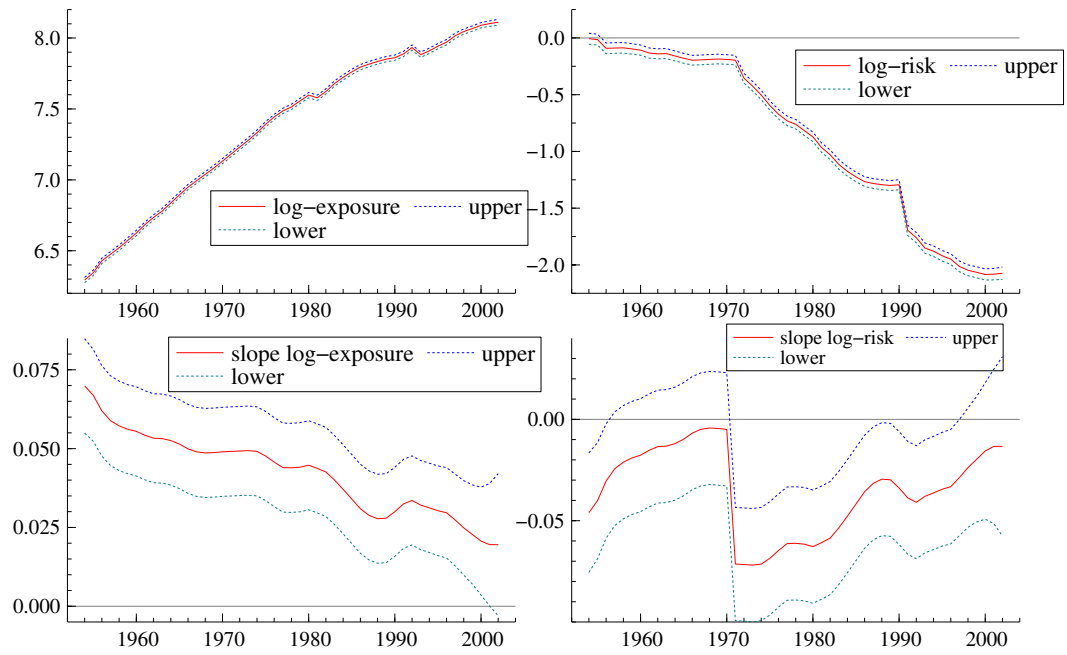


Figure 2: Smoothed output for LRT applied to Victorian insurance data

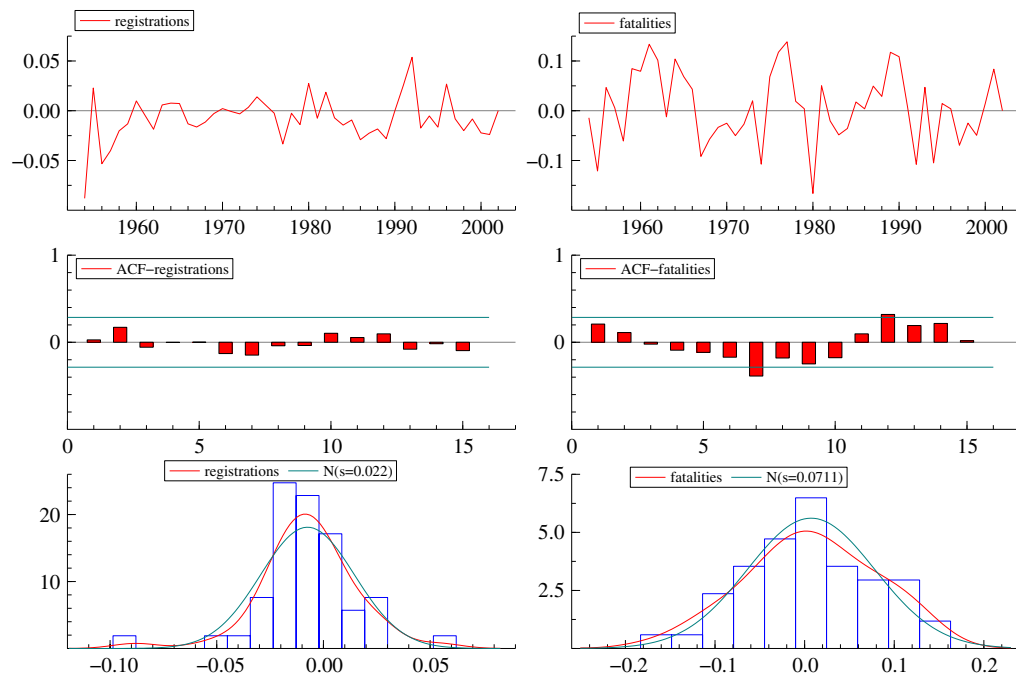


Figure 3: One-step-ahead residuals for the LRT applied to Victorian insurance data

fluctuations in Victorian GDP or unemployment. The model may account for such effects by the inclusion of a dynamic latent process or by explanatory variables.

References

- Allen, L. and A. Saunders (2003). A survey of cyclical effects in credit risk measurement models. *Technical report, BIS Working Paper 126*.
- BIS (2004). Basel II: International convergence of capital measurement and capital standards: A revised framework. Technical Report 107, Bank of International Settlements.
- de Jong, P. and P. Boyle (1983). Monitoring mortality a state-space approach. *Journal of Econometrics* 23, 131–146.
- Dominici, F., A. McDermott, and T. Hastie (2004). Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association* 99, 938–948.
- Fahrmeir, L. and S. Wagenpfeil (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risk models. *Journal of the American Statistical Association* 91, 1584–1594.
- Finkenstadt, B. and B. Grenfell (2000). Time series modelling of childhood diseases: A dynamical systems approach. *Applied Statistics* 49, 187–205.
- Gaudry, M. (1984). Drag, un modèle de la demande routière, des accidents et leur gravité, appliqué au québec de 1956–1986. Technical Report Publication CRT-359, Centre de recherche sur les Transports, et Cahier #8432, Département de sciences économiques, Université de Montréal.
- Gaudry, M. and S. Lassarre (Eds.) (2000). *Structural Road Accident Models: The International DRAG Family*. Oxford: Elsevier Science Ltd.
- Harvey, A. and C. Fernandes (1989). Time series models for insurance claims. *Journal of the Institute of Actuaries* 116, 513–528.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Ledolter, J., S. Klugman, and C. Lee (1990). Credibility models with time-varying trend components. *ASTIN Bulletin* 21.
- Morton, A. and B. Finkenstadt (2005). Discrete time modelling of disease incidence time series by using markov chain monte carlo methods. *Applied Statistics* 54, 575–594.
- Oppe, S. (1989). Macroscopic models for traffic and traffic safety. *Accident Analysis and Prevention* 21, 225–232.
- WHO (2004). World report on road traffic injury prevention. Edited by Margie Peden, Richard Scur .eld, David Sleet, Dinesh Mohan, Adnan A. Hyder, Eva Jarawan and Colin Mathers.