

*Macroscopic Road Crash Models with Time-Varying
Parameters*

TRB Annual General Meeting January 2006

Phillip Gould

pgould@feweb.vu.nl

Monash University Accident Research Centre

Overview

- A very (very) brief overview of time-varying parameter unobserved components models.
- Introduce a more structured time-series model of accident exposure, risk and severity — the latent risk time series model (LRT).
- Apply the model to Australian fatality data.
- Use the model to benchmark the performance of different Australian states.

This work is the result of a collaboration with Siem Jan Koopman (Free University), Frits Bijleveld and Jacques Commandeur (SWOV).

Background

- Macroscopic models are generally applied at country or state-wide level.
- Over time we may expect relationships to change — e.g. relationship between traffic volume and accidents.
- Reasons for the change may be due to technological innovation, policy changes, demographic shifts etc...
- Time-variation may be deterministic (very loosely “predictable”) or **stochastic** (random).

Time-Varying Parameters — Unobserved Components

Observed time series (e.g. monthly fatalities) can be decomposed into a number of unobserved components (UCs), including:

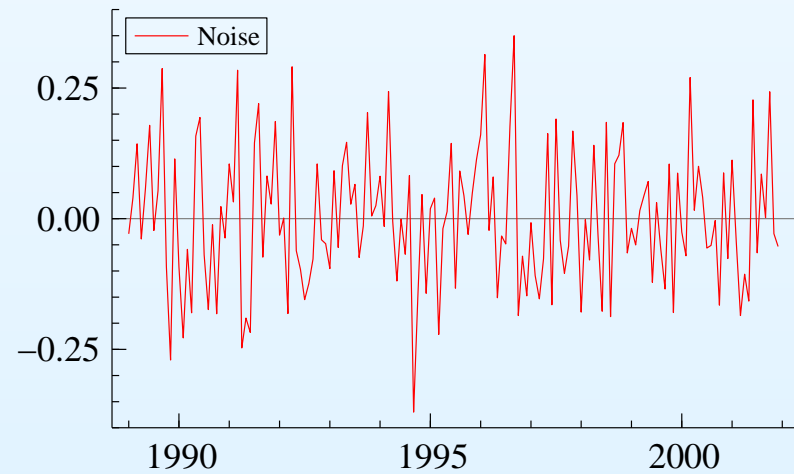
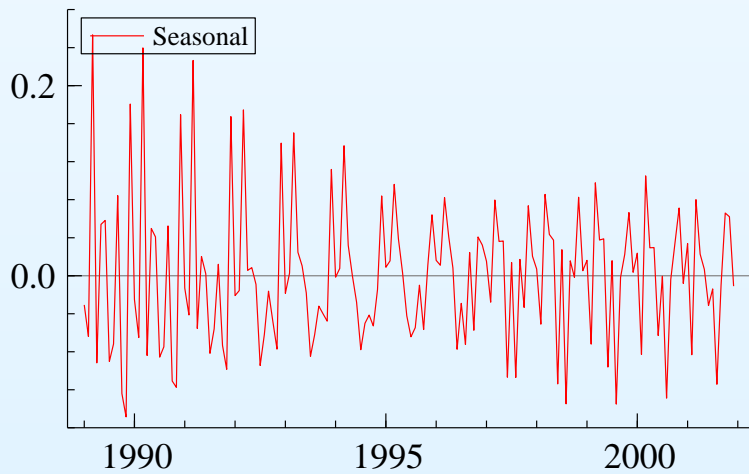
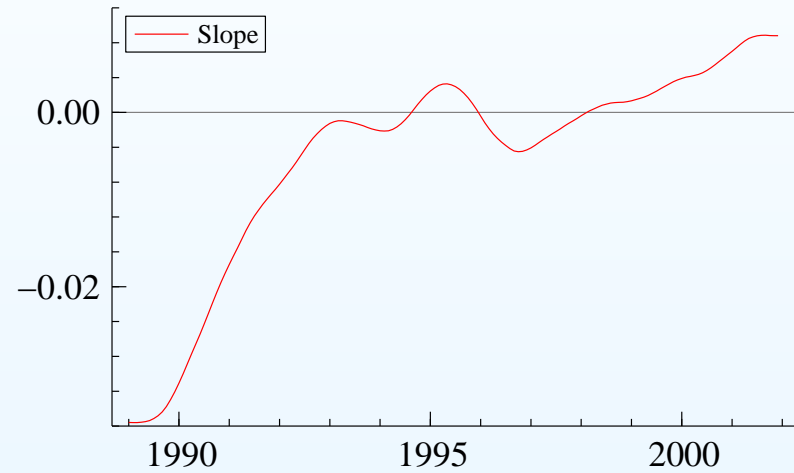
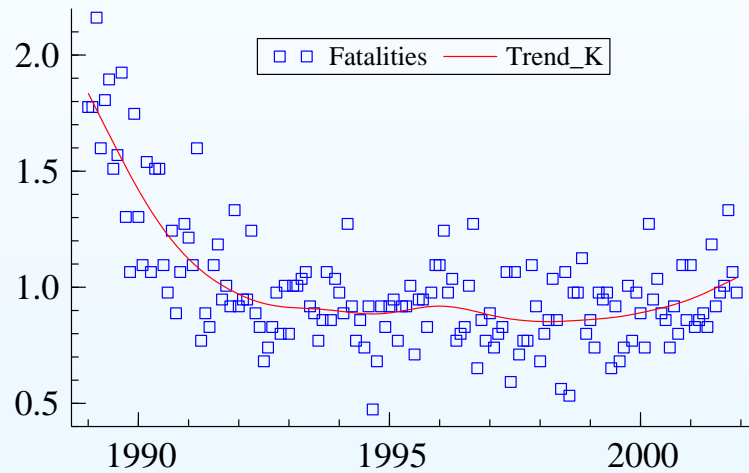
- level (trend, long-run average)
- slope (expected change in the long-run average)
- seasonal (e.g. monthly data)
- noise (random fluctuations)
- cycles (sinusoidal)
- regression variables

e.g.

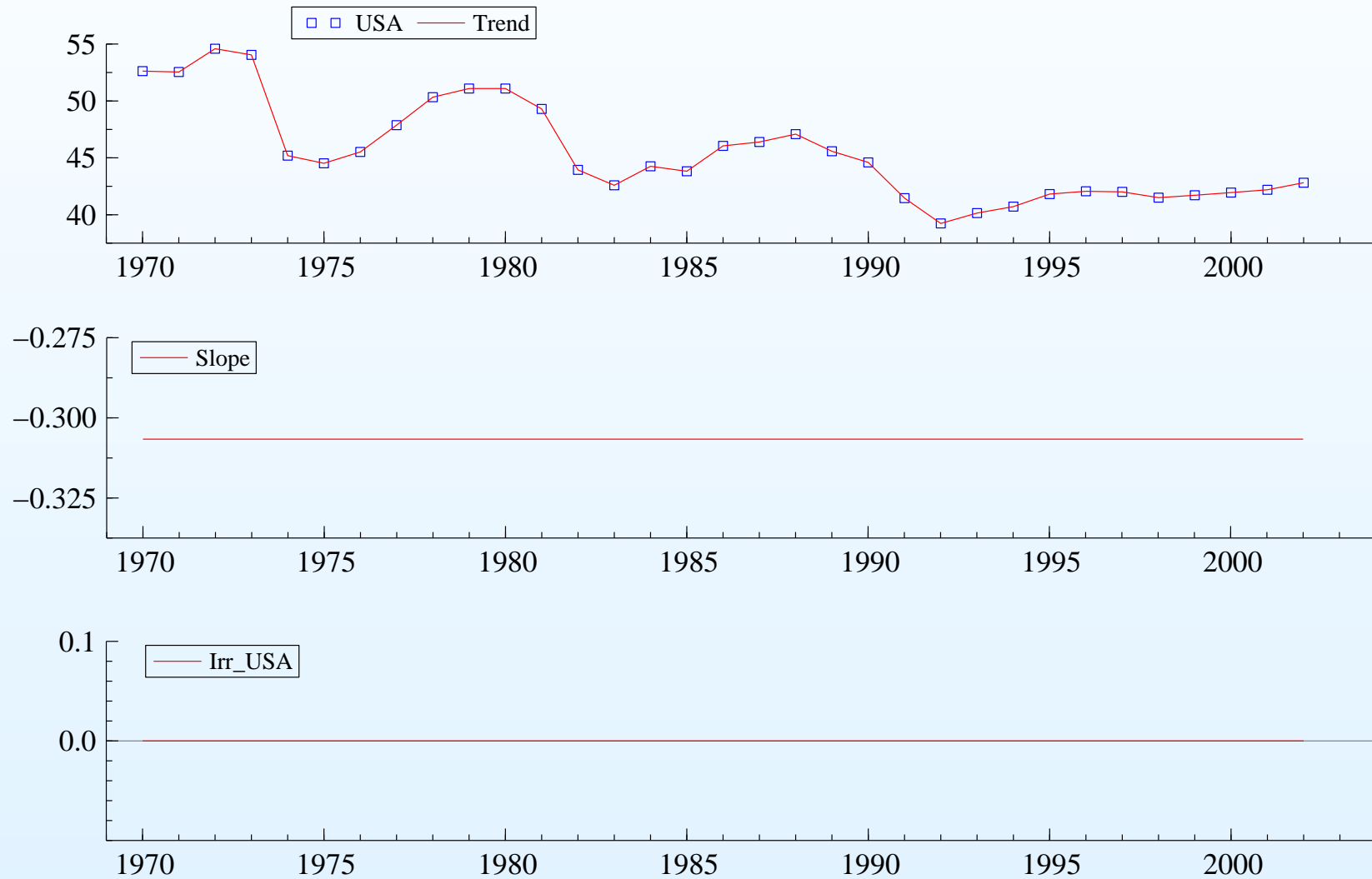
$$observed_t = level_t + slope_t + seasonal_t + noise_t$$

Estimation via the Kalman filter (STAMP) or exponential smoothing (PhiCast).

Fatalities in Victoria, Australia



Fatalities in the USA



Adding Structure — Defining Risk

There are many types of risk

1. financial risk due to changes in asset value (VaR calculations)
2. accounting risk (audit)
3. the risk of an event occurring (e.g. **road accidents**, company defaults, insurance claims)

Our focus is on item 3 — “event risk”, which is a common interest to many fields of research

- actuaries, credit risk managers
- **road safety researchers**
- **other transport applications** (e.g. risk of lateness)
- epidemiologists (e.g. risk of contracting a disease)

Time Series Models of Event Risk

$$\text{Loss}_t = \text{Exposure}_t \times \text{Risk}_t \times \text{Severity}_t \times \text{Error}_t$$

Relatively few articles in the academic literature.

- de Jong and Boyle (1983) — mortality rates using a state space approach
- Harvey and Fernandes (1989) — a state space model for insurance claims
- Oppe (1989) — a model for road accident risk
- Gaudry (1984) — the DRAG model for road accident exposure, risk and severity
- some recent epidemiological applications

The Latent Risk Time Series Model (LRT)

The 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 $i = 1, \dots, k_z$.

The model is written as

$$x_{it} = E_{it} \times U_{it}^{(x)},$$

$$y_{it} = E_{it} \times R_{it} \times U_{it}^{(y)},$$

$$z_{it} = E_{it} \times R_{it} \times S_{it} \times U_{it}^{(z)},$$

where E_{it} , R_{it} , and S_{it} are the latent variables exposure, risk and severity for group i at time t , respectively. $U_{it}^{(x,y,x)}$ are the multiplicative error terms.

The LRT in Additive Form

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

$$\log x_{it} = \mu_{it}^{(E)} + \varepsilon_{it}^{(x)},$$

$$\log y_{it} = \mu_{it}^{(E)} + \mu_{it}^{(R)} + \varepsilon_{it}^{(y)},$$

$$\log z_{it} = \mu_{it}^{(E)} + \mu_{it}^{(R)} + \mu_{it}^{(S)} + \varepsilon_{it}^{(z)},$$

where

- $\mu_{it}^{(E)} = \log E_{it}$ is log-exposure
- $\mu_{it}^{(R)} = \log R_{it}$ is log-risk
- $\mu_{it}^{(S)} = \log S_{it}$ is log-severity for group i and time t .

They are all *unobserved* or *latent* processes.

Latent Processes

The latent processes evolve according to local linear trend (LLT) specifications.

The LLT model for the log-exposure factor (for example) is given by

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

While the error terms are independent over time, there may be correlation between errors for different equations:

- common trends and slopes
- common components for risk, exposure and severity

No suitable “point and click” software, so models were coded using Ox and SsfPack.

Estimation via the *Kalman filter*

The LRT Model in State Space

The LRT model can be expressed in a form suitable for Kalman filtering:

$$\begin{pmatrix} \log x_t \\ \log y_t \\ \log z_t \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \eta_t, \quad \delta_t = \delta_{t-1} + \zeta_t,$$

where

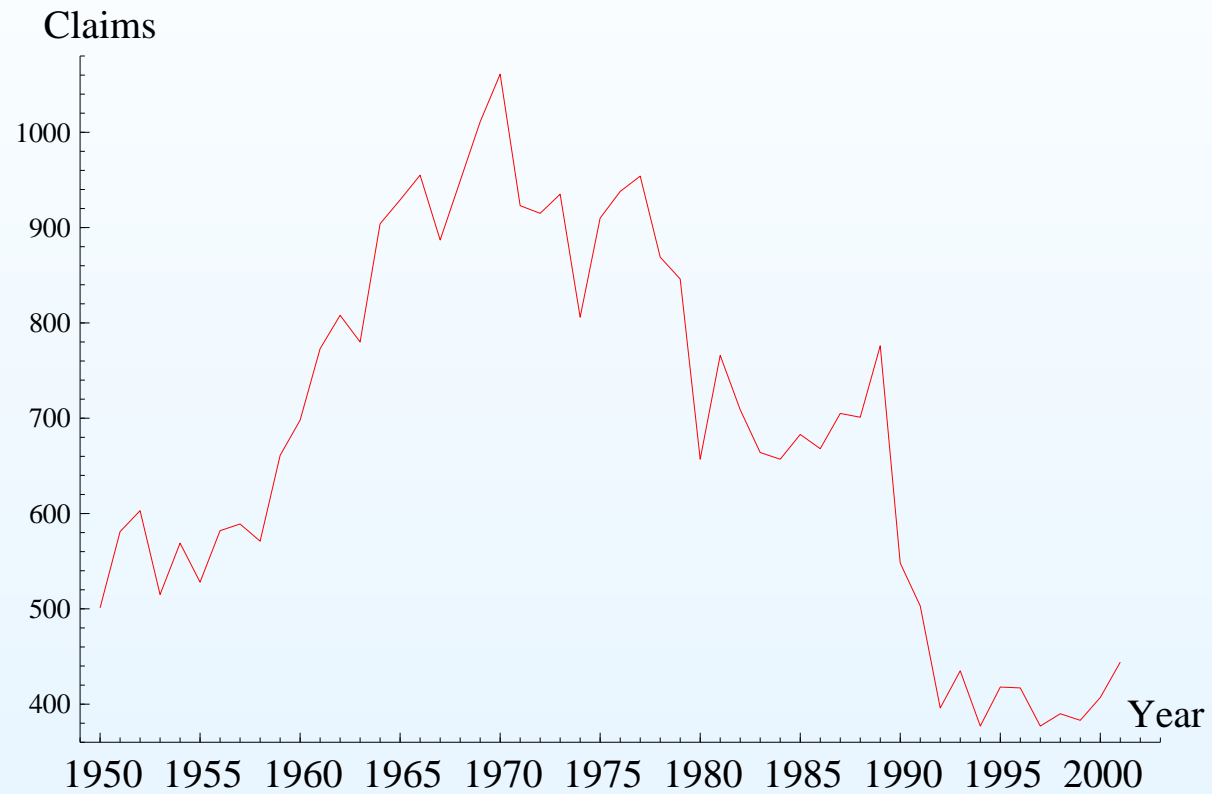
$$\mu_t = \begin{pmatrix} \mu_t^{(E)} \\ \mu_t^{(R)} \\ \mu_t^{(S)} \end{pmatrix}, \quad \delta_t = \begin{pmatrix} \delta_t^{(E)} \\ \delta_t^{(R)} \\ \delta_t^{(S)} \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_t^{(E)} \\ \eta_t^{(R)} \\ \eta_t^{(S)} \end{pmatrix},$$

$$\zeta_t = \begin{pmatrix} \zeta_t^{(E)} \\ \zeta_t^{(R)} \\ \zeta_t^{(S)} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_t^{(E)} \\ \varepsilon_t^{(R)} \\ \varepsilon_t^{(S)} \end{pmatrix}.$$

Application: Analysis of Victorian Fatalities

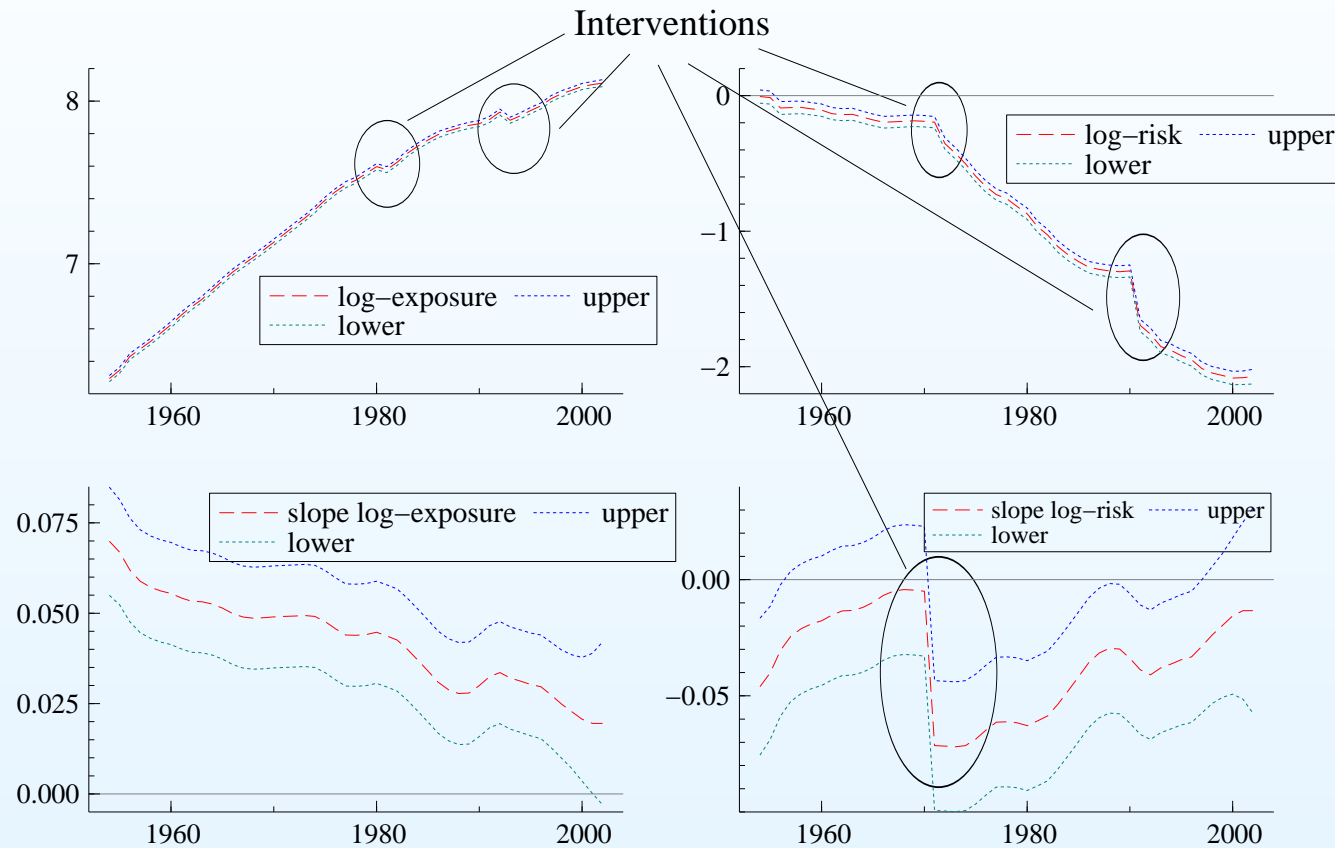
- The state of Victoria, Australia has a single, compulsory third-party motor vehicle insurer
- We study claims made on fatalities — the number of claims is the number of fatalities y_t
- The volume of exposure is measured by vehicle registrations x_t
- We use annual data from 1950 – 2001
- Aim is to measure and explain past evolution of exposure and risk
- Interventions are used at suitable times to explain changes in the system — introduction of seat belt laws, advertising campaigns etc...

Victorian Fatalities



Annual Victorian road crash fatalities

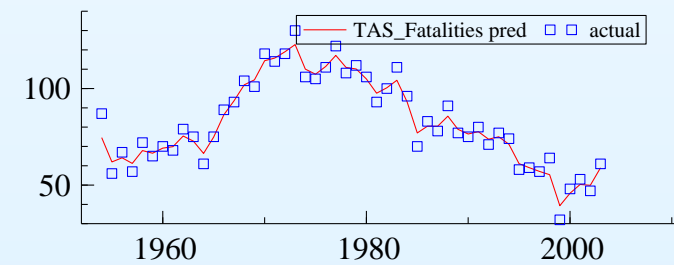
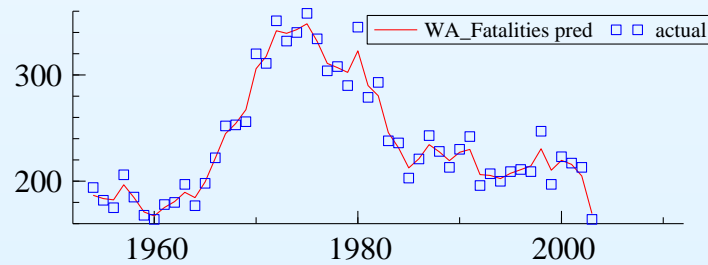
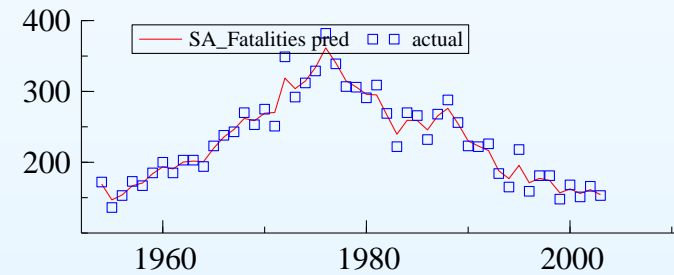
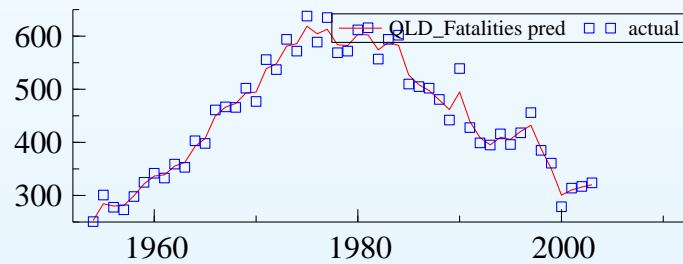
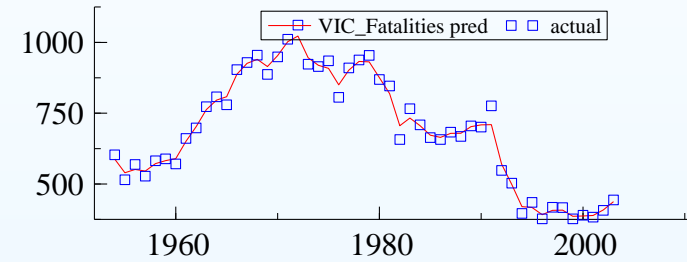
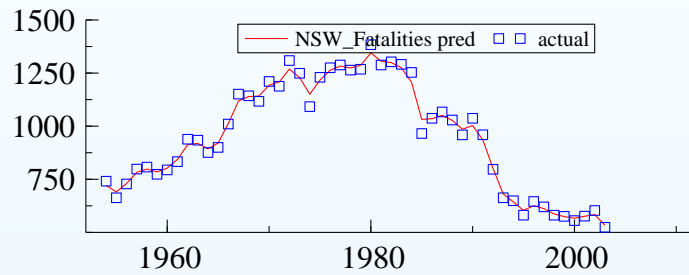
LRT Output: Smoothed States



- All 5 interventions are significant
- Large correlations between Exposure and Risk — common trends

Multiple Cohorts

The model was also run across the six Australian states.



Full Multivariate Model

Correlations for the level component (loosely “medium-term”)

Correlations	NSW	VIC	QLD	SA	WA	TAS
NSW	1.00	0.64	0.81	0.33	0.91	0.92
VIC	0.64	1.00	0.07	0.94	0.26	0.88
QLD	0.81	0.07	1.00	-0.28	0.98	0.53
SA	0.33	0.94	-0.28	1.00	-0.10	0.67
WA	0.91	0.26	0.98	-0.10	1.00	0.68
TAS	0.92	0.88	0.53	0.67	0.68	1.00
Eigenvalues	3.91	2.09	0.00	0.00	-0.00	0.00
Eigenvectors	1.00	-0.19	0.02	-0.99	-0.85	-0.43
	0.77	0.74	1.00	0.33	0.26	-0.37
	0.71	-0.81	0.50	0.27	-0.05	1.00
	0.49	1.00	-0.49	0.57	-0.64	0.47
	0.84	-0.65	-0.45	1.00	0.02	-0.69
	0.99	0.25	-0.53	-0.58	1.00	0.35

Common Factor Model

The common factor LRT (FLRT) model is written as:

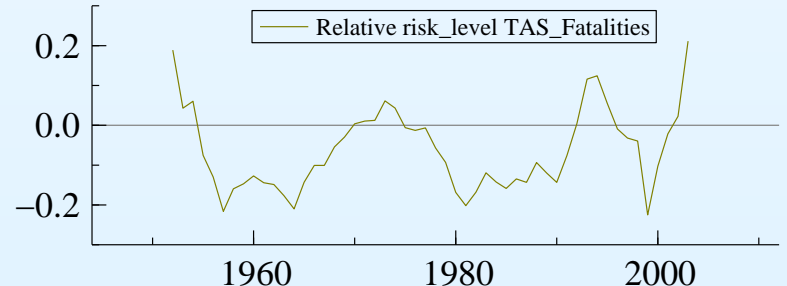
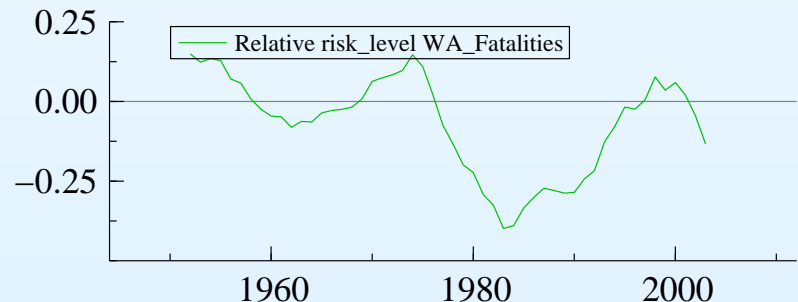
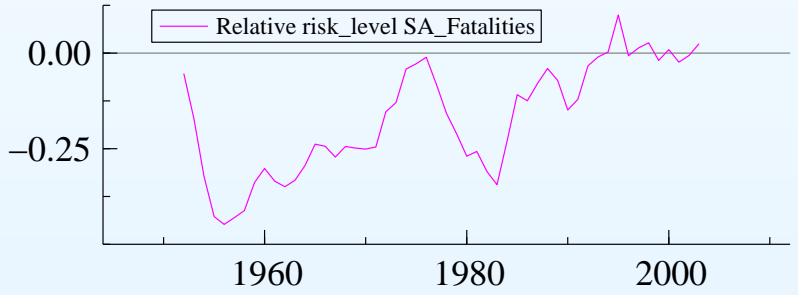
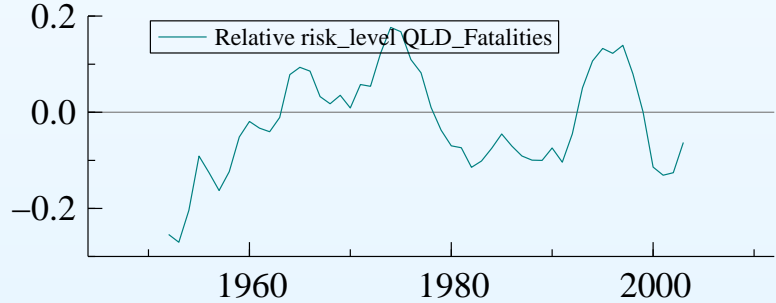
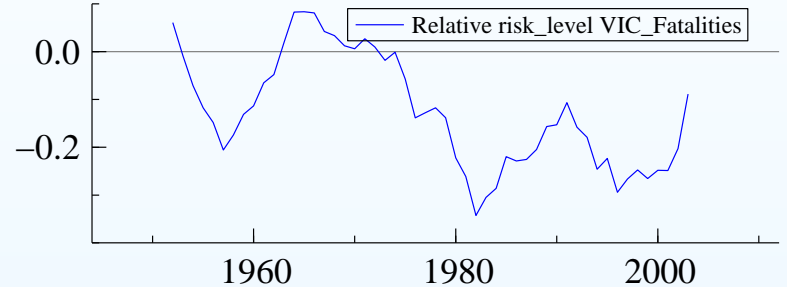
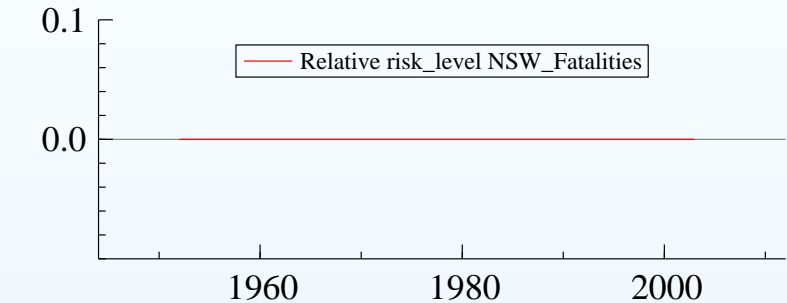
$$\begin{aligned}\log x_{it} &= \mu_{it}^{(E)} + \lambda_{it}^{(E)} \mu_t^{(CE)} + \varepsilon_{it}^{(x)} \\ \log y_{it} &= \mu_{it}^{(E)} + \lambda_{it}^{(E)} \mu_t^{(CE)} + \mu_{it}^{(R)} + \lambda_{it}^{(R)} \mu_t^{(CR)} + \varepsilon_{it}^{(y)},\end{aligned}$$

where $\lambda_{it}^{(E)}$ and $\lambda_{it}^{(R)}$ are the i th loadings on the common factors for risk ($\mu_t^{(CE)}$) and exposure ($\mu_t^{(CR)}$).

We define relative risk by

$$\mu_{it}^{RR} = \mu_{it}^{(R)} + (\lambda_{it}^{(R)} - 1) \mu_t^{(CR)}.$$

Common Factor Model — Relative Risk



Comparing the States

	Average Rel. Risk	Final Rel. Risk	Final Slope Rel. Risk
NSW	0	0	0
VIC	-0.1276	-0.1488	-0.0014
QLD	-0.0226	-0.1016	-0.0124
SA	-0.1692	0.0122	0.0064
WA	-0.0699	-0.1306	-0.0041
TAS	-0.0419	0.0319	0.0006

Conclusions

The LRT model...

- uses existing time series methods
- can be used on a variety of road crash or social problems (e.g. testing the effect on severity of new vehicle safety measures)
- can be used at an aggregate or disaggregate level (across regions, road-user type and time)
- allows us to benchmark safety performance across different economies
- provides good forecasts relative to other macro models (not demonstrated in this presentation)