Basing fatality forecasts on the joint development of mobility and road safety

Jacques J.F. Commandeur (SWOV, VU Amsterdam), Sylvain Lassarre (IFSTTAR)

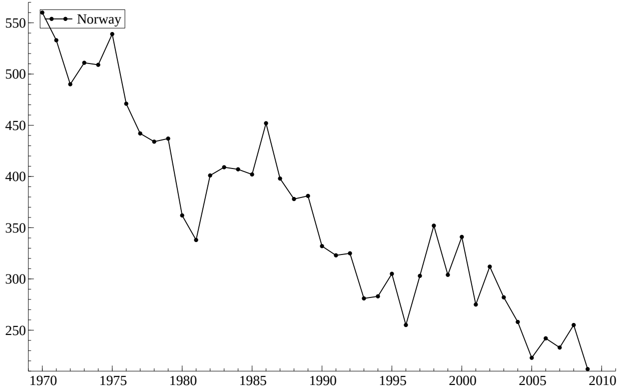
November 22, 2012

Overview

- Univariate structural time series models
- Bivariate models
  - SUTSE model
  - LRT, the Latent Risk model
- Dependencies between fatalities and exposure
- Model choices
  - LLT, the Local Linear Trend model
  - LRT, the Latent Risk models

What is a time series?

- A time series is the result of the repeated measurement of one and the same phenomenon.
- Example: Road traffic fatalities in Norway 1970-2009:



Objectives of time series models

- The objectives of time series models are to:
  - obtain an adequate *description* of a time series by establishing the *trend* in the series
  - find *explanations* for the observed developments
  - obtain *forecasts* of developments of a series into the (unknown) future
- Proper forecasts can only be obtained if the trend in a time series has been appropriately captured.

What's so special about time series?

- Unlike cross-sectional data, successive observations in a time series are usually *not independent*
- For example: chances are quite small that the number of fatalities next year will be completely different from the number of fatalities this year
- We have two types of univariate structural time series model:
  - deterministic linear trend models
  - stochastic linear trend models

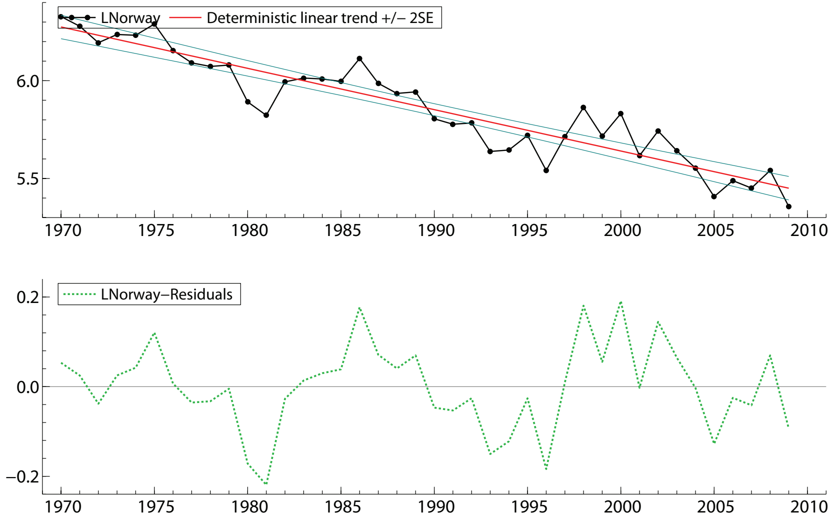
The deterministic linear trend model

- The deterministic linear trend model for obtaining a description of the trend in a time series y_t of annual data is

$$log(y_t) = a + bt + e_t, \qquad e_t \sim \mathrm{NID}\left(\sigma_e^2\right)$$

  where $t = 1, \ldots, n$ and n is the number of time points in the series, and the predictor variable $t = 1, 2, \ldots, n$ is *time itself*.
- Proper statistical conclusions from a trend model can only be derived if the errors or residuals e_t are normally and *independently* distributed.
- Note: both the intercept a and the slope b are treated deterministically, that is, are not allowed to change over time.

Results deterministic linear trend model



The stochastic linear trend model

- Subjecting both the intercept a and the slope b to a *random walk* yields the stochastic (local) linear trend model:

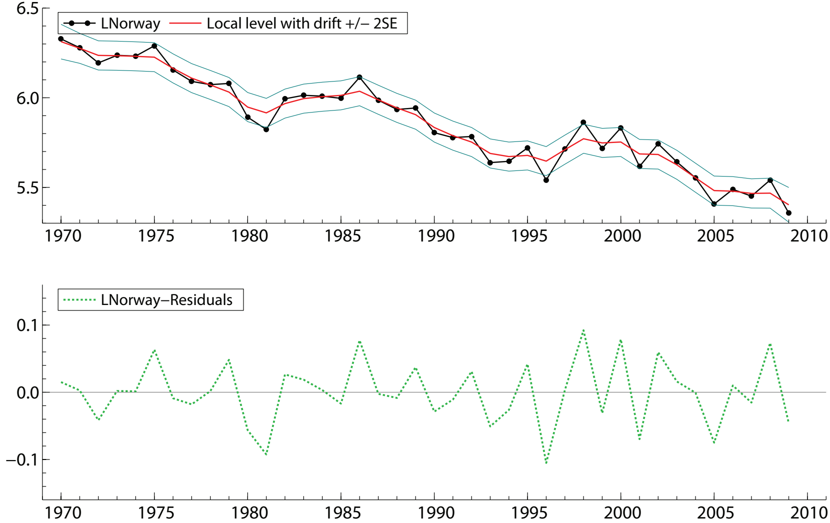
$$
\begin{aligned}
log(y_t) =& a_t + e_t, \qquad & e_t \sim \text{NID}\,(\sigma_e^2) \\
a_{t+1} =& a_t + b_t + \xi_t, \qquad & \xi_t \sim \text{NID}\,(\sigma_\xi^2) \\
b_{t+1} =& b_t + \zeta_t, \qquad & \zeta_t \sim \text{NID}\,(\sigma_\zeta^2)
\end{aligned}
$$

- Note: both the intercept or level component and the slope component are now treated stochastically, that is, they are allowed to change over time.

The stochastic linear trend model

- Special cases are:
  - The level a is allowed to change over time, but not the slope b: the local level model with drift;
  - The slope b is allowed to change over time, but not the level a: the smooth trend model.

Results of the local level model with drift



So which model is best for a given time series?

- This is decided by inspection of
  - the values of the variances,
  - the diagnostic tests for independence and normality of the residuals,
  - the fit of the model (using the Akaike Information Criterion).
- For series of fatalities we find different types of model to be adequate, depending upon the series at hand.
- For series of exposure data we often find the smooth trend model to be the most appropriate.

Types of forecast

- From deterministic linear trend models, forecasts just continue the straight line based on *all years* in the series, with a confidence interval that is usually much too tight, giving a false sense of certainty.
- From stochastic linear trend models, forecasts continue the level and slope mainly based on the *years at the end* of the series, with a confidence interval that becomes wider and wider as time proceeds, as is to be expected on intuitive grounds.

Interventions

- To all these models, interventions can be added to evaluate the effects of road safety measures:

$$log(y_t) = a_t + \lambda w_t + e_t, \qquad e_t \sim \text{NID}\left(\sigma_e^2\right)$$

where w_t is a dummy variable containing zeroes before, and ones at and after the road safety measure was introduced; λ is an unknown regression coefficient.

Example: Fatalities in France 1975-2010

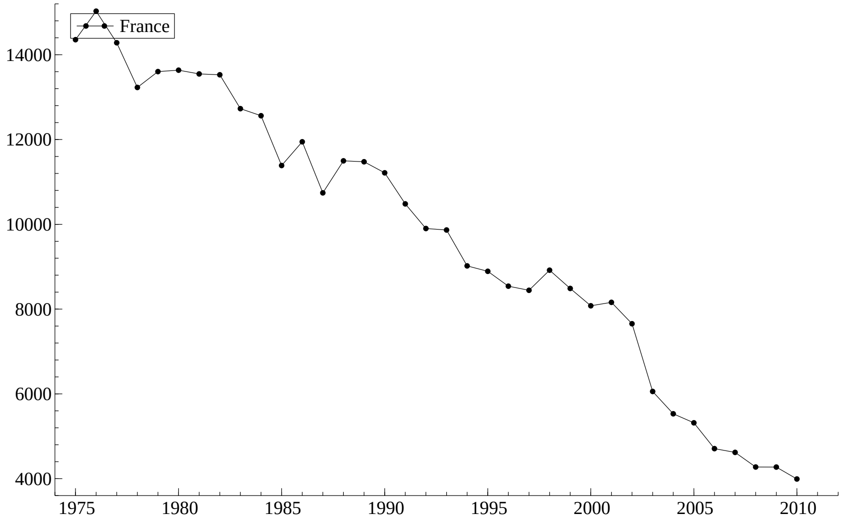
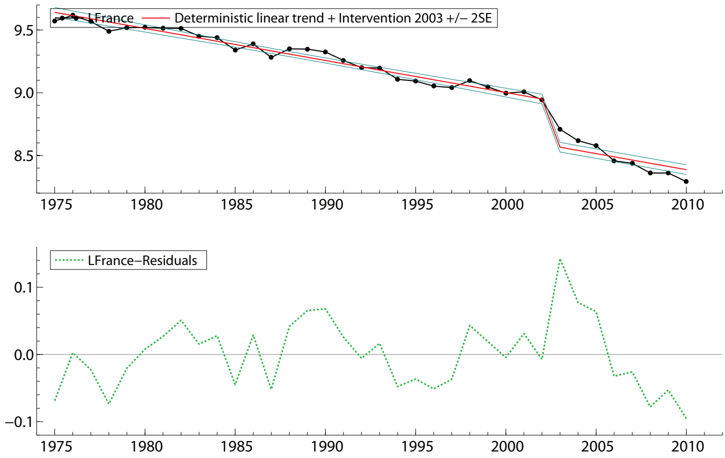
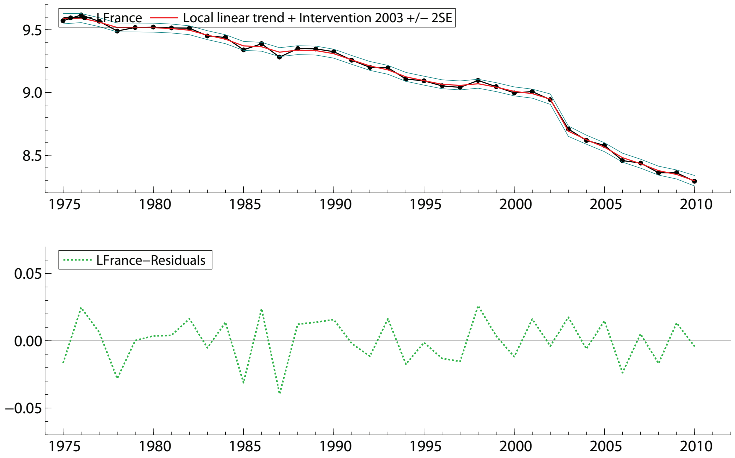


Illustration of an intervention in France, deterministic trend



- The effect of the intervention is estimated to be -0.359 (a 30% drop) with a t-value of -11.65.

Illustration of an intervention in France: local linear trend



- The effect of the intervention is now estimated to be -0.216 (a 19% drop) with a t-value of -4.84.