Time series analysis: from econometrics to epidemiology

Antonio Gasparrini

London School of Hygiene and Tropical Medicine

29 September 2010

Time series: definition

A time series is defined as a collection of observations sampled at **equally-spaced** and **ordered** time points

Statistically, the series is treated as a sequence of n random variables $Y_1, \ldots, Y_t, \ldots, Y_n$, assumed to be a single realization of a *discrete-time stochastic process* $\{Y_t\}$

ARIMA models

Many probabilistic models are based on the assumption of *(weakly) stationarity* of the series:

- Constant $\mu = \mathrm{E}(Y)$
- $\mathrm{Cov}(Y_t, Y_s) = \gamma(h)$, with $h = |t - s|$

Series usually exhibit *stochastic* or *deterministic* trends

Stationarity may be recovered by *auto-regressing*, *integrating*, and *averaging* (*filtering*) the series $\longrightarrow$ **ARIMA** models

Temporal decomposition

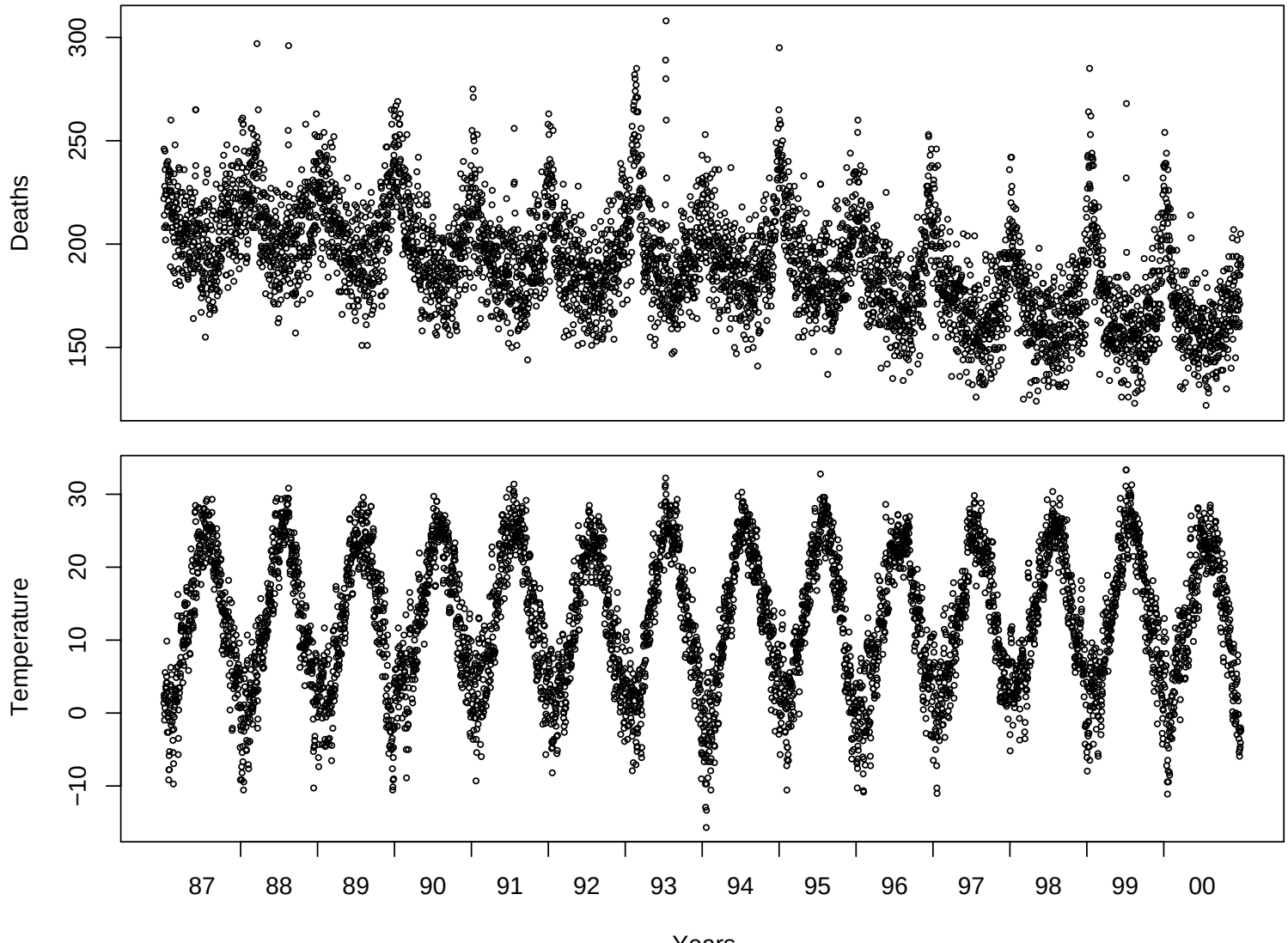
In epidemiological studies, the purpose of time series analysis is shifted from **prediction** to **estimation**

The series is decomposed into long-time and seasonal trends (or other components related to different timescales), and the contribution of **additional terms**

The series $\{Y_t\}$ is then described as the sum of *deterministic signal* plus a *stochastic stationary noise*

Temperature and mortality series

New York 1987-2000



Regression models

In modern applications, decomposition is performed through **regression models**

A general model to describe the series of observed outcomes y_t, with $t = 1, \ldots, n$ is given by:

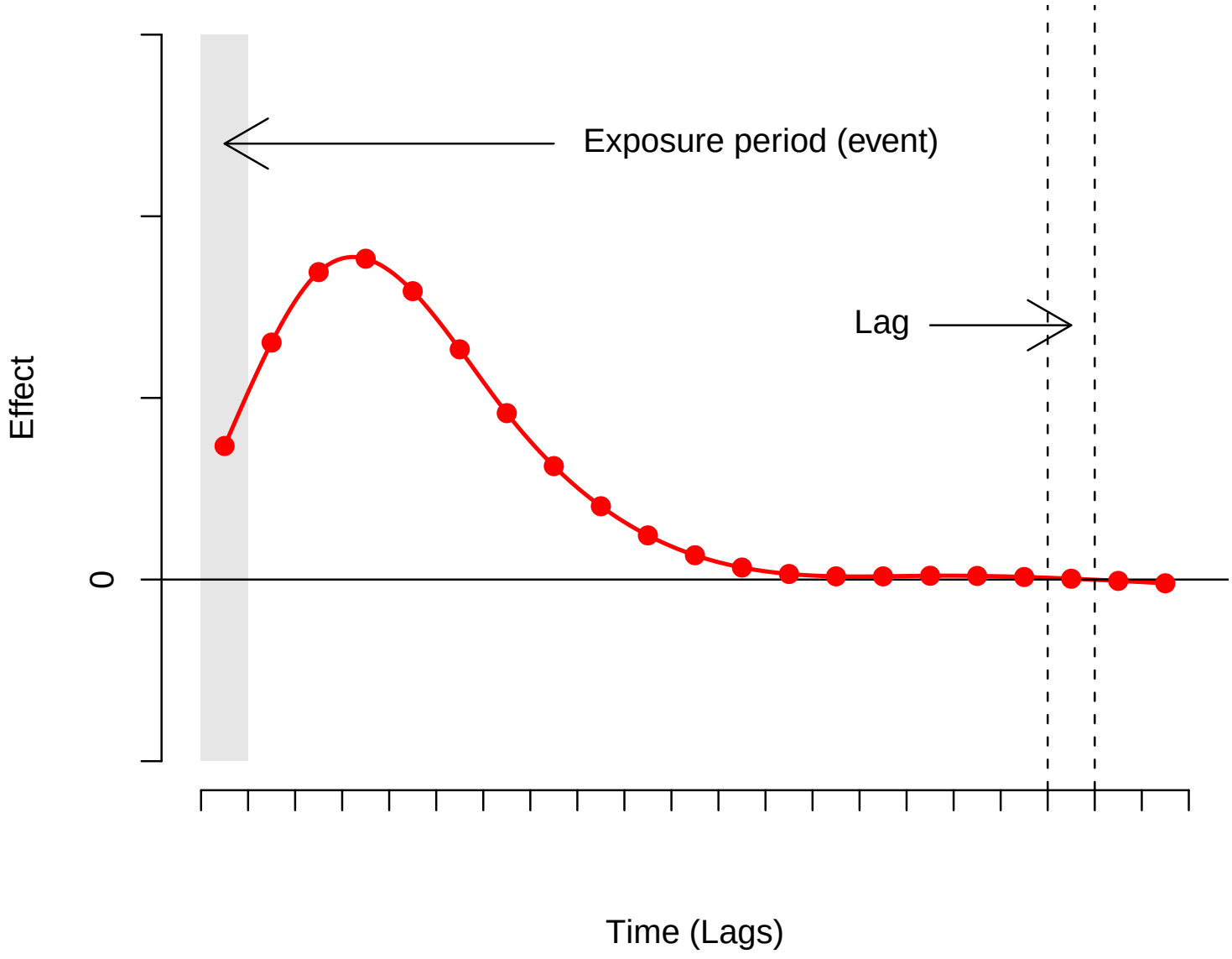
$$g(y_t) = \alpha + \sum_{j=1}^{J} s_j(x_{tj}; \beta_j) + \epsilon_t$$

Focus on the index t: **temporal structure** of the association

Methodological research topics

- Methods to describe the **temporal structure** of the association (delayed effects)
- **Smoothing techniques** for control of seasonality
- Methods to incorporate **residual correlation**
- **Interrupted time series**: before-after design

Delayed effects



Distributed lag (non-linear) models

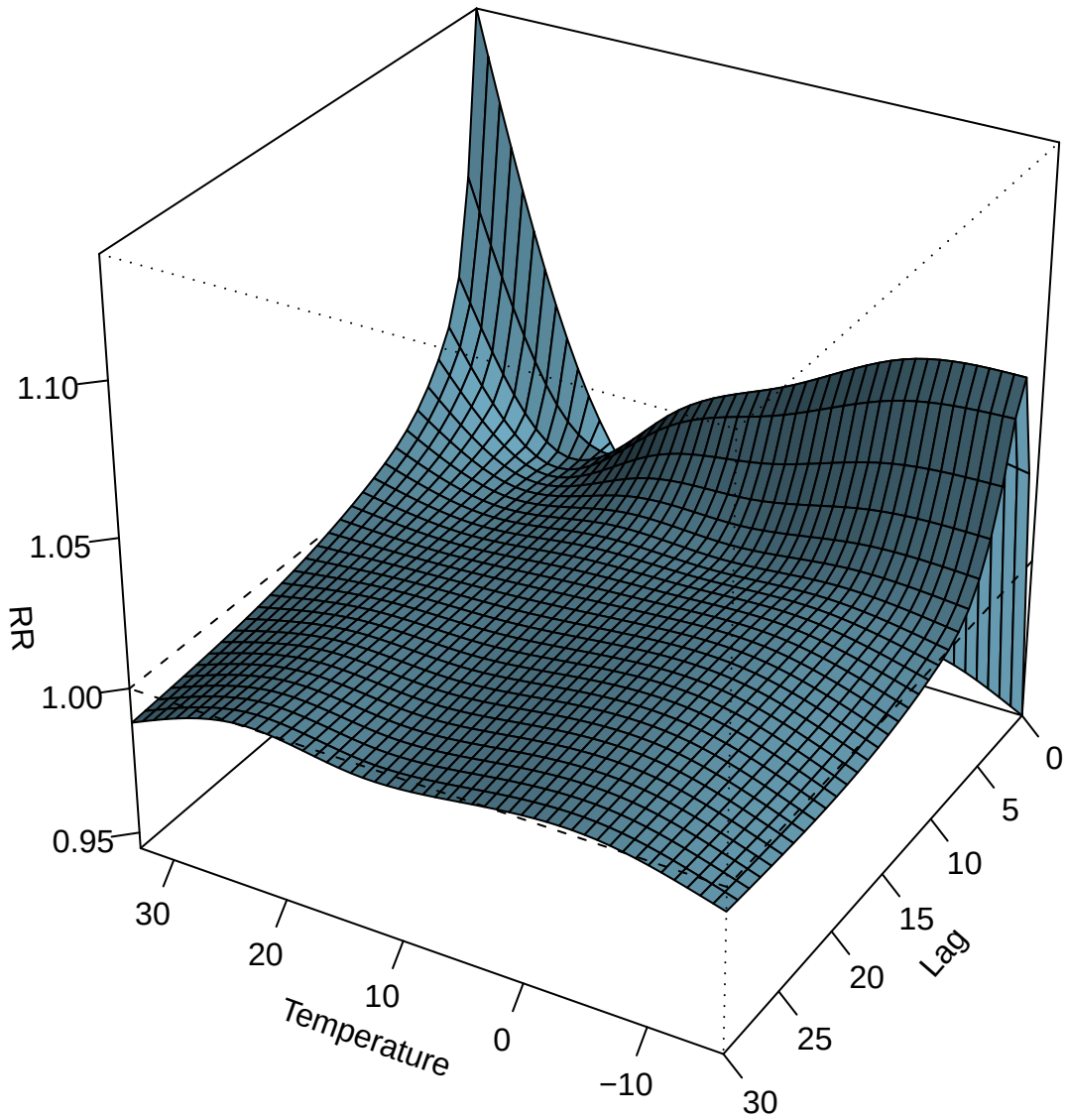
A statistical framework to describe simultaneously **non-linear** and **delayed** effects in time series data

DLNMs are expressed by the definition of a **cross-basis**:
bi-dimensional function describing the relationship along the spaces of predictor and lag

This framework is implemented in the R package `dlnm`

Temperature and mortality (I)

Chicago 1987-2000



Temperature and mortality (II)

Chicago 1987-2000