Time series regression: advancements in this new tool for epidemiological analyses.

Ben Armstrong
Antonio Gasparrini
LSHTM Feb 2012
Broad themes covered

Part 1 (Ben)
1. Introducing time series regression
2. Introducing distributed lag non-linear models (DLNMs)

Part 2 (Antonio)
1. Exploring patterns in DLNMs (& complex patterns generally) across “cities”
Time series regression – the context

$Y_i$

$x_{1,i}$

$x_{2,i}$

$t_i$
Time series regressions

**What:** (Poisson) regression with some time series elements

**Why:** Find acute effects of time-varying factors

**Why not:** Effect is chronic or exposure is time invariant

Statistical model needed to allow for:

- Measured time-varying risk factors (flu, pollution, ...)
- Overdispersion
- [Autocorrelation]
- **Unmeasured** smoothly varying risk factors ...
- **Lagged** effects (delay)
Model – general form

Quasi-Poisson regression model with:

\[ E(Y_i) = \exp\{f(t_i, \beta) + [\alpha + \sum_{k=1}^{m} \gamma_k x_{k,i} + \sum_{k=m+1}^{p} S_k(\lambda_k, x_{k,i}) + S_0(\lambda_0, i)]\} \]

\( t_i = \text{temperature on day } i \)

\( Y_i = \text{mortality count on day } i \)
Model – general form

\[ E(Y_i) = \exp\{ f(t_i, \beta) + [\alpha + \sum_{k=1}^{m} \gamma_k x_{k,i} + \sum_{k=m+1}^{p} S_k(\lambda_k, x_{k,i}) + S_0(\lambda_0, i)] \} \]

or \[ E(Y_i) = \exp\{ <\text{temperature effect}> + <\text{Other effects}> \} \]

\[ t_i = \text{temperature on day } i \]
\[ Y_i = \text{mortality count on day } i \]
Model – general form

\[ E(Y_i) = \exp\{f(t_i, \beta)\} \left[ \alpha + \sum_{k=1}^{m} \gamma_k x_{k,i} + \sum_{k=m+1}^{p} S_k(\lambda_k, x_{k,i}) + S_0(\lambda_0, i) \right] \]

or \[ E(Y_i) = \exp\{<\text{temperature effect}> + <\text{Other effects}>\} \]

\[ t_i = \text{temperature on day } i \]

\[ Y_i = \text{mortality count on day } i \]
... Control for smoothly varying risk factors

Options:
- **Explicitly seasonal**
  - Stratification: month (1-12) or week (1-52) indicators
  - Fourier (sine/cosine) functions
- **General smooth functions of time**
  - Stratification: month(1-72) or week (1-312)
  - Linear
  - Polynomial
  - **Cubic spline (illustrated above with 7 df/y)**
  - Non-parametric (GAM: LOWESS, Penalised spline ...)

---

The graph shows a scatter plot with dates on the x-axis and a measure of risk factors on the y-axis. The data points are spread across different dates, indicating variations in risk factors over time.
Impact of confounding

Unadjusted:

Adjusted:

• Time/season smooth curve ,
• Pollution
• Flu
• Day of week

[Air pollution benchmark model]

That’s all on confounder control
Functional form of temperature–mortality association

\( f(t_i, \beta) \)
Splines: eg natural cubic spline: NCS

1940’s

1990’s

\[ f(t_i, \beta) = \sum_{b=1,k} \beta_b t_{b,i} \]

\[ t_{b,i} \approx [(t_i - \{knot\}_b)^+]^3 \]

“Spline basis” variables

OK geeks – not quite right – but close enough
Var. of interest: Non-linear options

See additional practical for more detail
Lags

Lag = delay (in days) between exposure (heat or cold) and death
Exploring lagged effects 1:

• Use a linear threshold model
• Fit a **distributed lag model (DLM)** to the heat slope
DLM (heat): \[ \sum_{l=0}^{27} \gamma_p t_{heat,i-p} \]

London CVD 1976-1993
### Constrained Distributed lags (degrees above 20°C)

**Quartic polynomial**  
[Schwartz 2001]

<table>
<thead>
<tr>
<th>Lag</th>
<th>RR increment per degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>1.01</td>
</tr>
</tbody>
</table>

**Equation**

\[
DLM(heat) = \sum_{p=1}^{P} \gamma_p t_{heat,p,i}, \text{ where } t_{heat,p,i} = \sum_{l=0}^{L} l_p t_{heat,i-l}
\]

**Lag basis variable**
Distributed lag models and "harvesting"

Deficit – "harvesting"?
A summary of net effect from distributed lag models

Net RR increment (%) over 28 days per degree above 20 (95% CI):

5.0(3.7,6.3)

4.9(3.6,6.2)

4.9(3.7,6.2)
Distributed lag non-linear model

\( f(t_i, \beta) \) tweaked again
To capture curved associations with non-trivial lag dependence:

Need combined virtues of:

(a) Smooth temp-mort curves

(b) Smooth effects of lag.
Distributed lag non-linear models (DLNM)

Unrestricted choice of:

<table>
<thead>
<tr>
<th>Shape of temperature-mort graph</th>
<th>Shape of change with lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Step changes</td>
<td>• Step changes</td>
</tr>
<tr>
<td>• Thresholds</td>
<td>• Polynomial</td>
</tr>
<tr>
<td>• Cubic splines</td>
<td>• Cubic splines</td>
</tr>
<tr>
<td>• ...</td>
<td>• ...</td>
</tr>
</tbody>
</table>

Implemented in R package dlnm AG (and clutzy stata code - BA)
RR by temperature (NCS) and lag(NCS)
Temperature-mortality models: recap

\[ f(t_i, \beta) = \]

Thresholds: \[ \beta_{\text{cold}} t_{\text{cold}, i} + \beta_{\text{heat}} t_{\text{heat}, i} \]

Spline (gen): \[ \sum_{b=1}^{B} \beta_b t_{b,i} \]

DLM (heat): \[ \sum_{p=1}^{P} \gamma_p t_{\text{heat}, p,i}, \text{ where } t_{\text{heat}, p,i} = \sum_{l=0}^{L} l_p t_{\text{heat}, i-l} \]
Temperature-mortality models: +1

\[ f(t_i, \beta) = \]  

Thresholds : \( \beta_{\text{cold}} t_{\text{cold},i} + \beta_{\text{heat}} t_{\text{heat},i} \)

Spline (gen) : \( \sum_{b=1}^{B} \beta_b t_{b,i} \)

DLM (heat) : \( \sum_{p=1}^{P} \gamma_p \overline{t_{\text{heat},p,i}}, \text{where} \overline{t_{\text{heat},p,i}} = \sum_{l=0}^{L} l_p t_{\text{heat},i-l} \)

DLNM (gen) : \( \sum_{b=1}^{B} \sum_{p=1}^{P} \eta_{b,p} \overline{t_{b,p,i}}, \text{where} \overline{t_{b,p,i}} = \sum_{l=0}^{L} l_p t_{b,i-l} \)

Temperature basis  
(B df)  
+  
Lag basis  
(P df)  
=  
“Cross” basis  
(BXP df)

Armstrong 2006
RR by temperature (NCS) and lag(NCS)

5df temp X 4df lag = 20 df
Ways of looking at DLNMs

A: RR by temperature

B: RR by lag

RR vs 20°C

RR by temperature for given lag

RR by lag for given temperature
- Adding graph of total (28 day) effect
Part 1 wrap-up

• Time series regression allows investigation of acute effects; many applications:
  – Air pollution -> health outcomes
  – Weather -> health (incl. infectious diseases)
  – Circulating viruses -> mortality/admissions
  – Policy changes -> health (interrupted time series)
  – ....

• DLNMs allow flexible modelling of delayed effects with possibly non-linear exposure-response.