

Does air pollution modify the heat tolerance? estimating the threshold-line via segmented regression¹

*L'inquinamento dell'aria modifica la tolleranza al caldo? la stima della
'linea-di-soglia' attraverso la regressione 'segmented'*

Vito M.R. Muggeo, Mariangela Sciandra²
Dipartimento di Scienze Statistiche e Matematiche 'Vianelli',
Università di Palermo, e-mail: vmuggeo@dssm.unipa.it, sciandra@dssm.unipa.it

Riassunto: In questo lavoro viene presentato un modello in cui l'effetto della temperatura sulla mortalità è espresso attraverso una relazione 'segmented' in cui il punto di svolta che esprime la tolleranza al caldo non è costante ma dipende dai valori di un'altra esplicativa che misura l'inquinamento dell'aria.

Keywords: threshold, break-point, heat-tolerance, interaction effect

1. Introduction

Temperature and air pollution represent serious threats for human health in urban areas; short-term effects of such environmental factors are well known and documented by several studies carried out in the last decade in America and Europe as well. Typically a Poisson model is employed to analyze daily time series of deaths Y_t as a function of seasonality (for instance via a smoother $s(t)$), air pollution (PM10_t, linear term) and outdoor temperature TEMP_t:

$$\log E[Y_t] = s(t) + \delta \text{PM10}_t + \beta_1 \text{TEMP}_t + \beta_2 (\text{TEMP}_t - \psi)_+ \quad (1)$$

for $t = 1, 2, \dots, n$. The terms $\text{TEMP}_t + (\text{TEMP}_t - \psi)_+$ (where $(x - a)_+ = (x - a)I(x > a)$) account for the V-shaped relationship between mortality and temperature and the parameter ψ is the temperature value where mortality reaches its minimum; such value is understood to measure the heat tolerance since mortality increases as temperature goes beyond it. In order to investigate on possible interaction of temperature-by-pollutant on mortality, we assess whether the heat tolerance can depend on pollution. This implies that we put an index on ψ , that is replace the last term in the right hand side of equation (1) by:

$$\beta_2 (\text{TEMP}_t - \psi_t)_+ \quad \psi_t = (1 + \exp\{-(\kappa_0 + \kappa_1 \text{PM10}_t)\})^{-1} \quad t = 1, \dots, n.$$

We illustrate the model and relevant estimating algorithm in a more general context the next section.

¹Lavoro svolto con finanziamento MIUR 2006

²Address of correspondence: Parco d'Orleans, viale delle Scienze, edificio 13, 90128 Palermo

2. Methods and Application

In simple segmented regression the piece-wise relationship between the response and the variable Z , is modelled via the term $(Z - \psi)_+$ in the linear predictor; without losing in generality we can assume that the segmented variable Z is defined in the range $[0, 1]$ and thus the threshold ψ may be re-expressed via a logistic function:

$$\psi_i = \psi(\boldsymbol{\kappa}; x_i) = \frac{e^{\kappa_0 + \kappa_1 x_i}}{1 + e^{\kappa_0 + \kappa_1 x_i}} \quad i = 1, 2, \dots, n \quad (2)$$

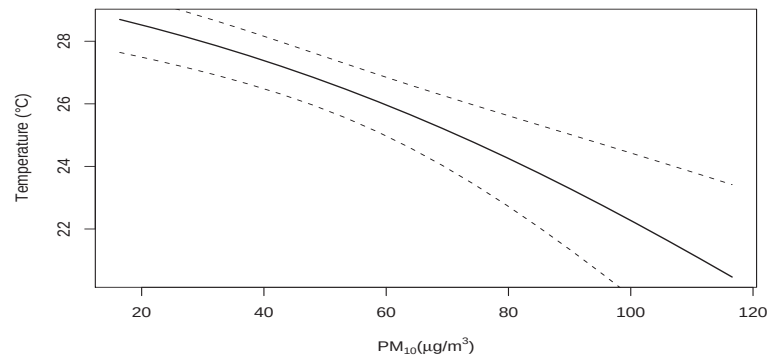
where $\boldsymbol{\kappa} = (\kappa_0, \kappa_1)^T$. Model (2) says that the threshold value for the variable Z is not constant but depends on the values of the variable X ; the resulting *threshold line* is constituted by the pairs (x_i, ψ_i) satisfying the (2).

The main advantage in using a logistic function is that the parameters $\boldsymbol{\kappa}$ are actually unbounded, making open the parameter space; in practice this implies that constrained optimization may be avoided. As an alternative to demanding quasi-Newton methods (Schwartz *et al.*, 1995), for estimation we propose a simple algorithm which generalizes the case of single breakpoint (Muggeo, 2003): the method relies on linearizing the non-linear term $f(\psi(\boldsymbol{\kappa})) = (Z - \psi(\boldsymbol{\kappa}))_+$, which is a composite function depending on two parameters via the logistic function (2). The procedure needs to evaluate at each iteration the first order Taylor expansion

$$f(\psi(\boldsymbol{\kappa})) \approx f(\psi(\boldsymbol{\kappa}^{(0)})) + (\boldsymbol{\kappa} - \boldsymbol{\kappa}^{(0)})^T \left. \frac{\partial f}{\partial \boldsymbol{\kappa}} \right|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}^{(0)}} \quad (3)$$

where the $(j + 1)$ th component of the gradient vector $\frac{\partial f}{\partial \boldsymbol{\kappa}}$ is the $(j + 1)$ th partial derivative. For the Palermo mortality data, the estimated threshold line displayed in Figure 1, is $\text{logit}(\hat{\psi}_t) = 2.2889 - 0.0165 \text{PM}_{10t}$ with standard errors of the relevant coefficients equal to 0.2109 and 0.0034 respectively.

Figure 1: Point estimate and 95% pointwise confidence intervals of the threshold-line temperature by PM_{10} .



References

- Muggeo V.M.R. (2003) Estimating regression models with unknown breakpoints, *Statist Med*, 22, 3055–3071.
- Schwartz F.P. *et al.* (1995) Threshold models for combination data from reproductive and developmental experiments, *J Am Statist Ass*, 90, 862–870.