Analysis of time series of respiratory disease and air pollution data

Márton Ispány[a], Juliana Bottoni de Souza[b], Valdério A. Reisen[b], Glaura C. Franco [b], Jane Meri Santos [b]

[a] Faculty of Informatics, University of Debrecen, Hungary
ispany.marton@inf.unideb.hu

[b] Departamento de Estatística, CCE-UFES, Departamento de Estatística, UFMG, Brazil

Abstract

Ambient air pollutants, such as PM_{10}, SO_2, NO_2, O_3 and CO, concentrations are considered to be explanatory covariates in the generalized linear model (GAM) for the analysis of hospital admissions due to respiratory diseases. Due the fact that the pollutant variables do not only possess serial dependence, but also interdependence amongst themselves, the hybrid model GAM-PCA-VAR is proposed which is the combination of the principal component analysis (PCA) and GAM along with a vector autoregressive (VAR) process. The PCA is used to eliminate the multicollinearity amongst the pollutants and the VAR model is used to handle the serial correlation of the data to produce white noise processes. For comparison purposes, two other models are also used in the empirical investigation: the GAM-PCA and GAM models. The first one uses the PCA components, computed from the original explanatory covariates in the GAM. The second one is the standard GAM using the original variables. The empirical results show that the GAM-PCA-VAR model is able to remove the autocorrelations from the principal components. In addition, this method produces more significant estimates of the relative risk (RR) for each pollutant, and generates better residual fits.

Keywords: GAM model; Principal component analysis; Vector autoregressive model; multicollinearity; relative risk.

MSC: 62M10; 62J12; 62H25; 62P12