A More Exhaustive View of Synthetic Environmental Indices [(1)]

Recupero di informazioni da indicatori sintetici ambientali

Francesca Bruno
Dipartimento di Scienze Statistiche "P.Fortunati", Università di Bologna
bruno@stat.unibo.it

Meri Raggi
DEIAgra, Università di Bologna
raggi@stat.unibo.it

Riassunto: Gli indicatori sintetici sono spesso utilizzati per riassumere situazioni complesse in un unico valore. Tuttavia, la sintesi in un solo valore fa perdere molte informazioni che potrebbero essere di grande utilità, soprattutto quando l'indice è di supporto al decisore pubblico. In questo lavoro si propone di utilizzare il modello logit multinomiale per recuperare informazioni da indicatori di qualità dell'aria utilizzando variabili meteorologiche come esplicatori.

Keywords: air-quality indices, multinomial logit analysis, relevant air pollutant, meteorological covariates.

1. Introduction

There is much debate regarding the use of reliable, portable, easy-to-compute and easily understood synthetic indices. Synthetic environmental indices aim to summarize a series of data using easy, clearly-defined rules. However, their use for the purposes of policy decision-making is not so straightforward due to the often hidden nature of initial quantities. This work examines the possibility of improving the information provided by synthetic environmental indices and of recovering information after the process of synthesis.

This paper considers air-pollution data from Bologna. Section 2 introduces the synthetic air-quality indices. In Section 3, a multinomial logit model is proposed and used to explain synthetic indices. We then briefly report some results and draw conclusions.

2. Air-Quality Indices

The present paper analyses an index contained within the general class of air-quality indices proposed in Bruno and Cocchi (2002). This class of indices can be constructed by reducing dimensions using successively aggregating functions.

[(1)]The present paper is financially supported by a 2004 grant (project n. 2004137478_001 sector: Economics and Statistics) for researches of national interest by the Italian Ministry of the University and Scientific and Technological Research.

In general, three aspects of air pollution need to be taken into consideration: time, space and type of pollutant. The first synthesis covers the time dimension, since monitoring sites continuously collect data. This leaves the other two dimensions: pollutants and monitoring sites. The most suitable order of aggregation has to be decided. The study by Bruno and Cocchi (2002) contains a full discussion of the two different orders of aggregation. Since the present work aims to offer an approach to cases of severe pollution, we have chosen maxima as aggregating functions, and we propose the first order of aggregation: monitoring sites-pollutants. The final index is thus:

$$I1_{R,M(M)} = \max_j f_R\left(X_{M\{j\}}\right) \text{ with } j=1,\ldots,J \tag{1}$$

where: j corresponds to the pollutant, $X_{M\{j\}} = \max_i\left(X_{q\{ij\}}\right)$ is the synthesis among I monitoring sites, and f_R is the standardizing function necessary when the aggregation is computed over various pollutants.

In the last few years, national laws for environmental standards have changed, and only one threshold is considered in air-quality evaluation. For this reason, since the final index value is taken from a continuous-scale final index, we consider the following transformation:

$$f_R(Y) = \frac{Y}{a_j} \cdot 100 \text{ with } j=1,\ldots,J. \tag{2}$$

where a_j represents the thresholds defined in Official Gazettes (G.U 87,13/04/2002; G.U 67/14, 9/3/2002). By using $I1_{R,M(M)}$, it is possible to synthesize the pollution in a large area, such as a city, using only one value. If this value exceeds 100, pollution is deemed severe. The higher the index, the more critical the pollution. Although the index offers a synthesis of the pollution situation, the information regarding which pollutant determines the value of the index gets lost during the process of aggregation, hence our proposed method of recovering this kind of information.

3. Data Analysis

3.1 The methodology

Multinomial logit regression is used (Hosmer and Lemeshow, 2000) to study the varying probability of categorical responses. This model considers the multi-choice case and its estimation eliminates problems of scale differences among choice alternatives. This model enables us to evaluate:

$$p_j = \frac{\exp(\beta_j^{'} x)}{\sum_j \exp(\beta_j^{'} x)} \qquad for\ j = 1,\ldots,J \tag{3}$$

where j corresponds to the possible categories, x represents the covariates and p_j is the probability of category j.

3.2 Bologna air-pollution data

The data used in this analysis are based on daily measurements of air pollution in the city of Bologna from 2000 to 2002. The pollutants considered are: particulate matter (PM_{10}), ozone (O_3) and benzene. Some meteorological variables (provided by the Emilia-Romagna Regional Agency for Environmental Protection) are also available. The multinomial logit model (3) is applied to this data-set, as this enables us to evaluate the probability of the most important pollutant determining the final index. Meteorological variables are used as covariates.

3.3 Some results

Table 1 contains the parameter estimates obtained analyzing all available data using a stepwise procedure and considering PM_{10} to be the reference category. The predictors selected by the procedure are: "Hmix": mixing height; "T_ave": average daily temperature; "Press": average daily air pressure; "Δpress": maximum daily variation in air pressure; and "Hol": a dummy variable indicating holidays. Moreover, two interactive factors - "Hmix*Press" and "Hmix* Δpress" - also emerged.

Table 1: *Multinomial logit regression: parameter estimates and p-values.*

	Covariate	Exp(β)	Sig.
Benzene	Hmix	0.756	0.007
	T_ave	0.900	0.000
	Press	0.816	0.000
	Δpress	1.318	0.000
	Hmix*Press	1.000	0.007
	Hmix* Δpress	1.000	0.003
O_3	Hol=0	0.140	0.000
	T_ave	1.077	0.015
	Press	0.818	0.005
	Δpress	1.289	0.015
	Hmix* Δpress	1.000	0.011

In Table 1, the Exp(β) column shows the variation in the odds of a one-unit change in the covariate. In particular, values greater than 1 indicate the positive effect of the meteorological covariate on pollutant occurrence probability in relation to the baseline pollutant. In the case of benzene, when we observe high mixing-height values, this means that the probability that benzene is the most relevant pollutant in the air pollution index is lower (proportional to 0.756) than is the probability that it is particulate matter. Average temperature and pressure produce separately a negative effect on the odds of the dominant pollutant being benzene rather than particulate matter, which means that when temperature and pressure are high, particulate matter is more influential than benzene. In the case of ozone (O_3), occurrence probability increases as average temperature and Δpress rise, by a multiplicative factor of 1.077 and 1.289 respectively (taking PM_{10} as the reference pollutant). Another relevant covariate is "Hol", working days (Hol=0) corresponding to a drop in O_3 occurrence probability (Exp(β)=0.140).

The selected model proves capable of correctly classifying 72% of those pollutants that determine air-pollution index values (Table 2). In particular, the model correctly classifies the occurrence of O_3 to a greater degree than it does the occurrence of other pollutants; this could be due to standard meteorological conditions encouraging high ozone values.

Table 2: *Multinomial Logit Regression: Classification Table.*

	Predicted			
Observed	Benzene	O_3	PM_{10}	**Percent correct**
Benzene	42	17	48	**39.3%**
O_3	4	191	23	**87.6%**
PM_{10}	24	34	153	**72.5%**
Overall percentage	13.1%	45.1%	41.8%	**72.0%**

Since the aim of air-quality indices is that of detecting and signalling severe episodes of air pollution, we decided to analyse these cases separately. Hence, we consider the 223 episodes of acute pollution as corresponding to index values higher than 100 (the severe threshold). In this case, the classification performance of the multinomial logit regression results is substantially better, since the model manages to correctly classify 89.5% of all cases (Table 3).

Table 3: *Multinomial Logit Regression: Classification Table (episodes of severe air pollution).*

	Predicted			
Observed	Benzene	O_3	PM_{10}	**Percent correct**
Benzene	10	0	7	**58.5%**
O_3	0	57	4	**93.4%**
PM_{10}	2	6	95	**92.2%**
Overall percentage	6.6%	34.8%	58.6%	**89.5%**

References

Bruno F., Cocchi D. (2002). A unified strategy for building simple air quality indices, *Environmetrics;* 13; 243-261.

G.U. n. 87 (13/04/2002) supplemento n. 77: Decreto Ministeriale 2 aprile 2002 n.60. Valori limite di qualità dell’aria ambiente per il biossido di zolfo, il biossido di azoto, gli ossidi di azoto, le particelle, il piombo, benzene e monossido di carbonio.

G.U. n. 67/14 (9/3/2002): Direttiva 2002/3/CE del 12/02/2002 relativa all’ozono nell’aria.

Hosmer, D. W., and Lemeshow S. (2000). *Applied Logistic Regression.* New York: John Wiley & Sons.