SIS 2017
Statistics and Data Science: new challenges, new generations

28–30 June 2017
Florence (Italy)

Proceedings of the Conference of the Italian Statistical Society

edited by
Alessandra Petrucci
Rosanna Verde

FIRENZE UNIVERSITY PRESS
2017

SIS 2017. Statistics and Data Science: new challenges, new generations : 28-30 June 2017 Florence (Italy) : proceedings of the Conference of the Italian Statistical Society / edited by Alessandra Petrucci, Rosanna Verde. – Firenze : Firenze University Press, 2017.
(Proceedings e report ; 114)

http://digital.casalini.it/9788864535210

ISBN 978-88-6453-521-0 (online)

Peer Review Process
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

Firenze University Press Editorial Board
A. Dolfi (Editor-in-Chief), M. Boddi, A. Bucelli, R. Casalbuoni, M. Garzaniti, M.C. Grisolia, P. Guarnieri, R. Lanfredini, A. Lenzi, P. Lo Nostro, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, G. Nigro, A. Perulli, M.C. Torricelli.

Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

SIS2017 Committees

Scientific Program Committee:
Rosanna Verde (chair), Università della Campania "Luigi Vanvitelli"
Maria Felice Arezzo, Sapienza Università di Roma
Antonino Mazzeo, Università di Napoli Federico II
Emanuele Baldacci, Eurostat
Pierpaolo Brutti, Sapienza Università di Roma
Marcello Chiodi, Università di Palermo
Corrado Crocetta, Università di Foggia
Giovanni De Luca, Università di Napoli Parthenope
Viviana Egidi, Sapienza Università di Roma
Giulio Ghellini, Università degli Studi di Siena
Ippoliti Luigi, Università di Chieti-Pescara "G. D'Annunzio"
Matteo Mazziotta, ISTAT
Lucia Paci, Università Cattolica del Sacro Cuore
Alessandra Petrucci, Università degli Studi di Firenze
Filomena Racioppi, Sapienza Università di Roma
Laura M. Sangalli, Politecnico di Milano
Bruno Scarpa, Università degli Studi di Padova
Cinzia Viroli, Università di Bologna

Local Organizing Committee:
Alessandra Petrucci (chair), Università degli Studi di Firenze
Gianni Betti, Università degli Studi di Siena
Fabrizio Cipollini, Università degli Studi di Firenze
Emanuela Dreassi, Università degli Studi di Firenze
Caterina Giusti, Università di Pisa
Leonardo Grilli, Università degli Studi di Firenze
Alessandra Mattei, Università degli Studi di Firenze
Elena Pirani, Università degli Studi di Firenze
Emilia Rocco, Università degli Studi di Firenze
Maria Cecilia Verri, Università degli Studi di Firenze

Supported by:
Università degli Studi di Firenze
Università di Pisa
Università degli Studi di Siena
ISTAT
Regione Toscana
Comune di Firenze
BITBANG srl

Index

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

A comparison between seasonality indices deployed in evaluating unimodal and bimodal patterns

Un confronto tra indici di stagionalità utilizzati nella misura di pattern unimodali e bimodali

Giovanni L. Lo Magno, Mauro Ferrante, Stefano De Cantis

Abstract This paper will discuss a recently proposed index for measuring seasonality, which is based on the solution of the well-known transportation problem. A specific characterization of the cost matrix will permit the taking into account of the cyclical structure of time periods, which characterizes the phenomenon under observation. Various features of the proposed index will be evaluated by comparing it with other indices which are commonly used in the measurement of seasonality, such as the Gini concentration index. Given the wide range of disciplines with an interest in the analysis of seasonal phenomena, the approach proposed may be of wide interest.

Abstract *Il presente lavoro discute un indice recentemente proposto per la misura della concentrazione stagionale e basato sulla soluzione del problema del trasporto. Una specifica caratterizzazione della matrice dei costi consente di tenere in considerazione la struttura ciclica dei periodi che caratterizza il fenomeno osservato. Diverse caratteristiche dell'indice proposto sono valutate confrontandolo con altri indici che sono comunemente utilizzati per la misura della stagionalità, come ad esempio l'indice di concentrazione di Gini. Considerata la varietà di ambiti interessati allo studio della stagionalità, l'approccio proposto può essere d'interesse da diverse prospettive.*

Key words: Seasonal amplitude, transportation problem, concentration index

Giovanni L. Lo Magno
Department of Economics, Business and Statistics, University of Palermo, e-mail: lomagno.gl@virgilio.it

Mauro Ferrante
Department of Culture and Society, University of Palermo, e-mail: mauro.ferrante@unipa.it

Stefano De Cantis
Department of Economics, Business and Statistics, University of Palermo, e-mail: stefano.decantis@unipa.it

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

1 Introduction

The notion of seasonality is very simple and very complex: referring to the former, it is a feature of many natural and human phenomena. The differing intensity of solar rays determines numerous effects on the environment, such as varying levels of humidity and temperature, but also on animal behaviour and human habits [2]. With reference to the latter, the notion of seasonality is an extremely complex concept, whose measurement and analysis is a challenge. Indeed, having reviewed the main indices used in different study contexts, we observed a lack of appropriate measures relating to seasonality, wich are capable of taking the cyclical structure of time periods into account. That is, the majority of indices currently used in measuring seasonality (e.g. the Gini concentration index, the coefficient of seasonal variation or the Theil index) do not take into account the natural ordering of time periods (e.g. months). Subsequently, given a pattern in which the total amount of the phenomenon of interest is concentrated in two consecutive months (e.g. January and February) and another pattern in which the phenomenon is concentrated in two very distant months (e.g. January and August), the currently-used indices would evaluate the two patterns as the same. One of the aims of the seasonality index discussed in this paper is to overcome this issue. Thus, after a discussion of a recently-proposed index for measuring seasonality, we will evaluate some of its properties in relation to how it behaves in evaluating unimodal and bimodal distributions.

2 A new index for measuring seasonality and its main properties

The approach we would like to propose for measuring seasonality is based on the solution of an appropriately defined transportation problem [1]. The transportation problem is a well-known, linear, minimization problem in which the goal is to minimize the cost of transferring units from a set of warehouses to a set of customers, satisfying the constraints given by the available resources and the requested demands.

We can define a *seasonal pattern* as the vector $\mathbf{P} = (y_1, y_2, \ldots, y_T)$, with observations for time periods from 1 to T, such that y_t, for $t = 1, 2, \ldots, T$, is non negative. The total amount of the observed phenomenon is $Y = \sum_{t=1}^{T} y_t$ and the average value is Y/T. Furthermore, let $\mathscr{A} = \{t : y_t > Y/T\}$ be the set of time periods for which the observed value is over the average; each of these time periods has a surplus $a_t = y_t - Y/T$. Similarly, let $\mathscr{B} = \{t : y_t < Y/T\}$ be the set of time periods with observed values under the average, each with deficit $b_t = Y/T - y_t$.

In order to eliminate seasonality, the T-dimensional pattern $\left(\frac{Y}{T}, \frac{Y}{T}, \ldots, \frac{Y}{T}\right)$ can be obtained by transferring units from time periods in $\mathscr{A}$ to time periods in $\mathscr{B}$. The amount which is transferred from time period $i \in \mathscr{A}$ to time period $j \in \mathscr{B}$ is x_{ij}. We can suppose that transferring one unit from i to j has a cost c_{ij}, thus the cost of all the transfers is $c = \sum_{i \in \mathscr{A}} \sum_{j \in \mathscr{B}} c_{ij} x_{ij}$. The minimum cost c^* which is

required to eliminate seasonality, through the aforementioned transfers, is the cost corresponding to the optimal solution to the following transportation problem:

$$\begin{array}{ll} \min c = \sum_{i\in\mathscr{A}} \sum_{j\in\mathscr{B}} c_{ij} x_{ij} & \\ \text{s.t.:} & \\ \sum_{j\in\mathscr{B}} x_{ij} = a_i, & \forall i \in \mathscr{A} \\ \sum_{i\in\mathscr{A}} x_{ij} = b_j, & \forall j \in \mathscr{B} \\ x_{ij} \geq 0, & \forall i \in \mathscr{A},\ j \in \mathscr{B} \end{array} \tag{1}$$

where the first set of constraints ensures that the amount a_i is transferred from each time period $i \in \mathscr{A}$; the second set of constraints ensures that each time period $j \in \mathscr{B}$ receives the amount b_j; and the third set of constraints ensures that each transfer is non-negative. The minimum cost c^* corresponding to the optimal solution to the transportation problem (1) is our absolute measure for seasonality, namely

$$S(\mathbf{P}) = c^* \tag{2}$$

In [3] we demonstrated that the maximum value of $S(\mathbf{P})$, holding Y constant, is:

$$S_{\max}(\mathbf{P}) = \frac{Y}{T} \max_{t\in M} \left\{ \sum_{j=1}^{T} c_{ij} \right\} \tag{3}$$

This result permits us to construct the relative seasonality index S_R which is bounded in the interval $[0,1]$:

$$S_R(\mathbf{P}) = \frac{S_R(\mathbf{P})}{S_{\max}(\mathbf{P})} \tag{4}$$

We recognize that the relation between time periods is cyclical. Time periods can be thought of as collocated in a circumference: the shorter arc with i and j as extremes is what is termed *cyclical distance*, and this is the unitary cost we consider for a transfer from i to j, namely:

$$c_{ij} = \begin{cases} |i-j| & \text{if } |i-j| \leq \dfrac{T}{2} \\ T - |i-j| & \text{otherwise} \end{cases} \tag{5}$$

Different unitary costs may be employed in the defining our seasonality index; however, when those costs defined in (5) are adopted, the S_R index has three important properties, which we deem desirable for measuring seasonality:

- *Scale invariance*: $S_R(\mathbf{P}) = S_R(\lambda\mathbf{P})$, with $\lambda > 0$.
- *Rotation invariance*: $S_R(y_1, y_2, \ldots, y_T) = S_R(y_T, y_1, y_2, \ldots, y_{T-1})$
- *Sensitivity to permutations*: generally, and excluding rotations (see the "rotation invariance" property), the permutations of a pattern are evaluated in a different way (the Gini index is permutation-invariant, thus it can not capture, for example, an increase in seasonality which is consequent on a reduction in the distance between two relevant modes).

3 Comparing seasonality indices

In this section we will compare: the Gini index G_C (normalized in the $[0,1]$ interval), the precipitation concentration index *PCD* [5] and the relative seasonality index S_R, when evaluating unimodal and bimodal patterns. Unimodal and bimodal patterns, obtained by adapting von Mises distributions to the discrete case, were used in these comparisons.

The von Mises distribution [4] is a circular, continuous, unimodal and symmetric distribution, with support in the $[0,2\pi]$ interval. Its shape depends on two parameters: the first is the expected value μ, which corresponds to the mode; the second is $\kappa \geqslant 0$ and it affects the degree of concentration. The higher is κ, the greater is the concentration around the mode; if $\kappa = 0$, then the von Mises distribution coincides with the circular uniform distribution.

In order to construct unimodal patterns of size T, we discretized von Mises distributions by partitioning their supports into T equal intervals of size $2\pi/T$ and then calculating the probability related to each interval. Thus, the value y_i in each resulting pattern is the probability which is related to the i-th interval of the von Mises' support and, consequently, $Y = 1$. In this study 241 von Mises distributions were considered for the unimodal patterns; these distributions have the same μ, but $k = 0, 0.1, 0.2, \ldots, 24$. The parameter μ was set to π/T, namely the midpoint of the first interval. Fig. 1 shows the results of all the indices which were applied to the 241 unimodal distributions under consideration. In order to clarify how the shape of the distribution changes as κ increases, radar charts for the distributions with $\kappa = 0, 4, 8, 12, 16, 20, 24$ have been included in the lower part of Fig. 1. It can be observed that all the indices increase as κ increases, thus they correctly capture the concentration parameter κ. Furthermore, the relation between the three indices is $S_R < G_C < PCD$, with the only exception for the distribution with $\kappa = 0$, for which all the indices have the value 0.

In order to observe the behaviour of the indices when they evaluate bimodal patterns, size 12 bimodal patterns were constructed as discrete versions of a 50% mixture of two von Mises distributions, both with the same κ parameter value. The distance d between the two modes varied between 0 (the unimodal case) and 6 (the maximum possible distance between two time periods). 1 and 5 were considered as two extreme values for κ, which corresponded to a low and a high concentration case around the two modes respectively. The indices values for the patterns when $\kappa = 1$ are plotted on Fig. 2 and those for the patterns when $\kappa = 5$ are displayed in Fig. 3.

Fig. 2 demonstrates that, for bimodal patterns where $\kappa = 1$, the indices decrease in value while d increases; furthermore they exhibit similar values. However, Fig. 3 also shows that the indices decrease in value while d increases, but they decrease at a quite different rate. To understand why this difference, it has to be noted that if κ is very high then there is a high concentration around the two modes, thus patterns where $d > 2$ can be approximately expressed as $\mathbf{P} = (y_1 = 0.5, 0, \ldots, 0, y_{1+d} = 0.5, 0, \ldots, 0)$. Thus, for $d > 2$, all the patterns can be roughly considered as permutations. As the Gini index is relatively insensitive to permutations, its value is quite

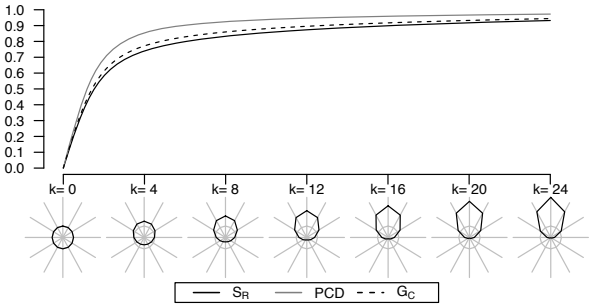

Fig. 1 Relative seasonality index S_R, corrected Gini index G_C and precipitation concentration degree index *PCD*, all calculated for discretized von Mises distributions with various values of κ and $T = 12$. Radar charts are reported for the distributions with $\kappa = 0, 4, 8, 12, 16, 20, 24$. The regular dodecagon in each radar chart corresponds to the uniform pattern.

stable for patterns where $d > 2$. This is not a desirable behaviour because a seasonality index should significantly decrease when the distance between two highly concentrated modes increases. In contrast, the *PCD* index clearly reveals decreasing values but it is zero when $d = 6$. This occurs because *PCD* equals zero for patterns where the first half of the pattern is the same as the second, like the pattern $(1, 3, 5, 1, 3, 5)$. A seasonality index should not return zero for patterns like these. Thus, the only index which shows a desirable behaviour is S_R, which decreases at an appreciable rate and does not reach zero when $d = 6$.

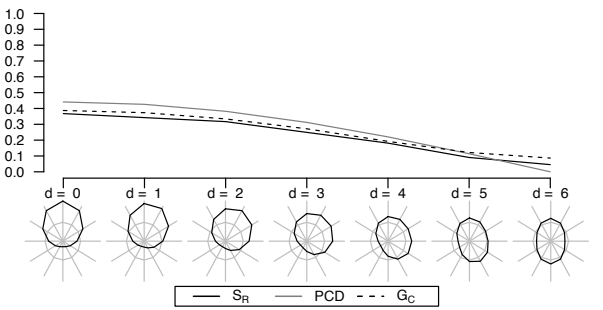

Fig. 2 Relative seasonality index S_R, corrected Gini index G_C and precipitation concentration degree index *PCD*, all calculated for discretized, bimodal von Mises distributions with $T = 12$, $\kappa = 1$ and values of the distance between the two modes $d = 0, 1, \ldots, 6$. Radar charts are reported for the distributions with $\kappa = 0, 4, 8, 12, 16, 20, 24$. The regular dodecagon in each radar chart corresponds to the uniform pattern.

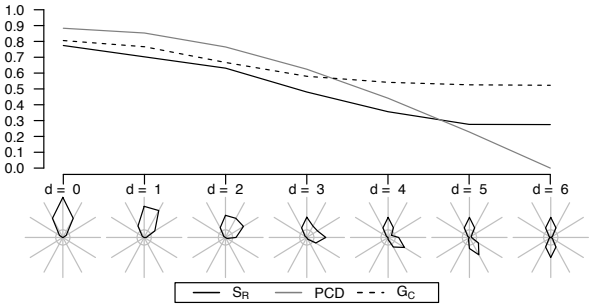

Fig. 3 Relative seasonality index S_R, corrected Gini index G_C and precipitation concentration degree index *PCD*, all calculated for discretized, bimodal von Mises distributions with $T = 12$, $\kappa = 5$ and values of the distance between the two modes $d = 0, 1, \ldots, 6$. Radar charts are reported for the distributions with $\kappa = 0, 4, 8, 12, 16, 20, 24$. The regular dodecagon in each radar chart corresponds to the uniform pattern.

4 Concluding remarks

The authors of this paper contend that the challenge of measuring seasonality has not received adequate attention in the literature. For example, the Gini index does not take into account the natural ordering of time periods. In addition to permitting a specification of the cost matrix which takes into account for the cyclical structure of time periods, our seasonality index performed well when applied to unimodal and bimodal patterns. Although a set of desirable properties has been highlighted in this paper, we hope that new realated issues will come to light as a result of this research. Given these challenges and considering the interdisciplinary nature of seasonal phenomena, the topic of measuring seasonality merits greater attention and from a wider range of points of view.

References

1. Dantzig, G. B.: Linear Programming and Extensions. Princeton University Press, Princeton (1963)
2. De Cantis, S., Ferrante, M.: Seasonality and tourism. The Sage Encyclopedia of Travel and Tourism. Sage (2016)
3. Lo Magno, G.L., Ferrante, M., De Cantis, S.: A new index for measuring seasonality: a transportation cost approach. (*Submitted to Mathematical Social Sciences*).
4. Jammalamadaka, S.R., SenGupta, A.: Topics in circular statistics. Series on multivariate analysis; vol.5, World Scientific Publishing, Singapore (2001)
5. Zhang, L., Qian, Y.: Annual distribution features of precipitation in China and their interannual variations. Acta Meteorologica Sinica 17 (2), 146–163 (2003)