European Journal of Internal Medicine 55 (2018) e9

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Letter to the Editor

No correlation between health care expenditure and mortality in the European Union

ARTICLE INFO

Keywords:
Time series data
Correlation
Stationary data
Non-stationary data

The correlation analysis in time series has received considerable attention [1]. Its use plays an important role in the social sciences, finance and medical sciences. Spearman's or Pearson correlation coefficients are the most commonly used for estimation [1]. The assumption such as stationarity makes the validity of time series correlation coefficients [2].

The impact of aggregate health care expenditure on health outcomes (mortality) in European Union have been published as a letter to editor in 2016 by Lippi et al. [3]. The authors examined the correlation between aggregate health care expenditure and mortality in European Union from the year 2000 to 2013. Their results showed no correlation between aggregate health care expenditure and mortality in European Union [3]. However, the time trending of aggregate health care expenditure and mortality described by Lippi et al. [3] raises the question of non-stationarity (i.e. data statistical properties vary over time or random walk) that lead to spurious correlation of the results (See [1,2]). In spurious correlation, the data series appear to be correlated, but the correlations itself are meaningless [2]. For instance, the significance Spearman's correlation used by Lippi et al. [3]; comes from the time trend of the data that affect the results, thus misleading the interpretation. Getting correlations coefficients without checking the stationarity assumption lead to “nonsense correlation” [2]. The data is stationary if the statistical property of time series (e.g., the mean, variance, autocorrelation and the correlation coefficients) does not change over time [4]. Moreover, the normalized data used by Lippi et al. [3] showed the linear relationship (Spearman's correlation coefficient of 0.55, $p = 0.054$), originated from non-linear data from the year 2000 to 2013, still not a useful coefficient since the series are non-stationary (See [2]). The traditional normalization method applied by Lippi et al. [3] showing procapita expenditure ratio and death ratio include the minimum-maximum (where the data vary from 0, 1 to 2). This kind of normalization is not always applicable to time series data if the data are not stationary [5]. Normalization through the Z-score method could be used for the time series data to become stationary (i.e., mean, variance and autocorrelation are constant over time) (See [6]). Likewise, differencing (i.e. compute the differences between consecutive observations) of the data could be used to stabilize the variance of time series from the year 2000 to 2013 to achieve stationarity by eliminating time trend. Another look to identify stationarity is to check whether the data is stationary or non- stationary is the use of unit roots tests such as Augmented Dickey-Fuller (ADF) test, KPSS tests and Phillips-Perron test [7].

As a take home message for the readers is that the time series correlation between aggregate health care expenditure and mortality interpreted concisely by Lippi et al. [3] is non-stationary misleading the interpretation. Observing the correlations only make the scientists and philosophers to become worried about fallacies [8], and that's why economist perform further statistical test like Granger Causality Test to check causal links (i.e. [9]), knowing that correlations of variables alone are not enough to conclude the results [8]. Lastly, the failure to correct non-stationarity leads to spurious correlation results and falls of methodological concern.

Conflict of interest statement

All authors have no conflict of interest including any financial, personal or other relationships with other people or organization.

References

[1] Erdem O, Ceyhan E, Varli Y. A new correlation coefficient for bivariate time-series data. Phys. A Stat. Mech. Appl. 2014;414:274–84.
[2] Yule GU. Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. J R Stat Soc 1926;89(1):1–63.
[3] Lippi G, Mattiuzzi C, Cervellin G. No correlation between health care expenditure and mortality in the European Union. Eur J Intern Med 2016;32:e13–4.
[4] Yang K, Shahabi C. On the stationarity of multivariate time series for correlation-based data analysis. In data mining, fifth IEEE international conference on 2005. 2005.
[5] Ogasawara E, Martinez LC, De Oliveira D, Zimbrão G, Pappa GL, Mattoso M. Adaptive normalization: A novel data normalization approach for non-stationary time series. In Neural networks (IJCNN), the 2010 international joint conference on 2010. 2010. p. 1–8.
[6] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J. Comput. Sci. 2011;2(1):1–8.
[7] Gujarati Damodar. Basic econometrics. MC Graw-Hill International Edition 2009.
[8] Calude CS, Longo G. The deluge of spurious correlations in big data. Found Sci 2017;22(3):595–612.
[9] Byaro M, Kinyondo A, Musonda P. Economic growth and under-five malaria mortality in Tanzania mainland: from correlation analysis to causality. Int J Health 2017;5(1):91–6.

Mwoya Byaro[a,b,*], Abel Kinyondo[a]
[a] *Research for Policy Development, REPOA, 157 Mgombani Street, P. O. Box 33223, Dar es Salaam, Tanzania*
[b] *Zambia University, Lusaka, Zambia*
E-mail address: byaro2014@yahoo.com*E-mail addresses:* byaro2014@yahoo.com (M. Byaro), akinyondo@repoa.or.tz (A. Kinyondo)

[*] Corresponding author at: REPOA, 157 Mgombani Street, P. O. Box 33223, Dar es Salaam, Tanzania.

https://doi.org/10.1016/j.ejim.2018.04.021
Received 17 April 2018; Received in revised form 20 April 2018; Accepted 26 April 2018
Available online 22 May 2018
0953-6205/ © 2018 European Federation of Internal Medicine. Published by Elsevier B.V. All rights reserved.