Title
The impact of ultraviolet radiation on sunburn-related search activity

Journal
Dermatology Online Journal, 23(8)

Authors
Lospinoso, Danielle J
Lospinoso, Joshua A
Miletta, Nathanial R

Publication Date
2017

DOI
10.5070/D3238035998

Volume 23 Number 8 | August 2017
DOJ 23 (8): 2

The impact of ultraviolet radiation on sunburn-related search activity

Danielle J Lospinoso[1] DO, Joshua A Lospinoso[2] PhD, Nathanial R Miletta[1] MD

Affiliations: [1]Department of Dermatology – San Antonio Military Medical Center, San Antonio, Texas [2]Portia Statistical Consulting LLC, San Antonio, Texas

Corresponding Author: Danielle J. Lospinoso, Fort Belvoir Community Hospital, Department of Dermatology, 9300 DeWitt Loop, Fort Belvoir, VA 22060, Email: danielle@lospi.net

Abstract

We establish a strong, positive relationship between the Ultraviolet Index and Google search engine activity for sunburn-related terms in the United States. Using the Google Trends utility and data available from the National Weather service, we combine data from a twelve-year period to produce panel data for each state. We fit a time-series regression model of search activity and perform statistical tests on the resulting parameter estimates. This study lays the groundwork for using search-related data to assess the prevalence of, and attitudes about sunburn. By tracking the frequency of searches about preventative measures like "sunscreen" or "protective clothing" versus treatment measures like "sunburn relief," researchers could measure the effectiveness of awareness and prevention programs.

Keywords: internet, search, time-series analysis, sun protection, sunscreen, skin cancer prevention, sunburn

Introduction

Exposure to the sun is a major contributing factor in basal cell carcinoma, squamous cell carcinoma, and melanoma of the skin [1-4]. Risks associated with sun exposure can be mitigated through safe sun protective practices [5-8], but achieving behavioral change in affected populations continues to be a public health challenge [5, 9-15]. Better understanding patient mindset regarding sun exposure and the epidemiologic prevalence of sunburn is critical to designing effective preventative measures [6]. This study explores the utilization of end-user internet search engine activity as a mechanism to assess these variables. By combining ultraviolet index data from the National Weather Service [16] with search activity data from Google [17] over a twelve year period, this study demonstrates strong evidence that internet search engine activity is an effective and efficient tool for assessing both prevalence of and attitudes about sunburn in the United States.

The Google search engine records search activity for a variety of uses, one of which is the Google Trends utility [17]. Medical researchers have employed Google Trends to study patient populations within geographic locations in a number of settings. Research has established a relationship between variations in alloaerogens, such as pollens or spores, and Google search activity for associated symptoms [18-20]. Extensions facilitating collection of this data have been proposed that would facilitate analysis of Google Trends information in a broad spectrum of otolaryngological applications [21]. From an infectious epidemiologic perspective, researchers have already determined a strong link between search activity and influenza outbreaks [22, 23].

Sunburn is a form of a radiation burn that affects living tissue, including the skin [3]. The scientific evidence for causality between ultraviolet radiation and sunburn is overwhelming [4, 24-28]. To establish that sunburn-related search activity is strongly associated with prevalence of sunburn, the ultraviolet index reported by the National Weather Service can be used [16]. If there is a sufficiently strong statistical relationship between this index and sunburn-related search activity, it is proposed that search activity is an effective measurement of sunburn prevalence. Such a tool contains a wealth of metadata to include whether a population is searching for information about preventive measures, e.g. "best sunscreen" or about treatment, e.g. "peeling sunburn."

Findings

Google Trends (https://google.com/trends) was used to collect internet search frequency data for the term "sunburn." Time series were collected for all fifty states. Concatenated, these datasets formed 18,255 observations from January, 2004 to March, 2016. Depending on the raw frequency of traffic, Google Trends provides the time series in either weekly or monthly resolution. We refer to this aggregate collection of data as the sunburn search index (SSI).

The Climate Prediction Center (CPC) of the National Weather Service makes historical ultraviolet index (UVI) data available via a file transfer protocol server (ftp.cpc.ncep.noaa.gov). All fifty states had at least one measuring station available, with 1 states having more than one. CPC provides daily resolution for UVI at each station's location. Concatenated, these datasets formed 252,537 observations from January 2004 to December 2015. There are two versions of UVI available: clear sky and cloudy. The former ignores the dampening effect of cloud cover on UV radiation, whereas the second attempts to model this attenuation. Analysis was performed upon both versions of the index and the conclusions were identical. For simplicity, the cloudy UVI is presented for the remainder of the paper, as our statistical model selection indicated better fit with the cloudy variation.

In order to best analyze the datasets, they were joined. A join is the association of objects in one data source with objects that share a common attribute in another data source. All Google Trends search activity observations with weekly resolution were joined by averaging all observations within a given month. Next, the CPC's UVI data was converted to monthly resolution in a similar fashion, for all observations within a given state in a given month, and an average was calculated. With these modifications, the search activity and UVI could be joined at the state/month level. The resulting dataset ranged from January 2004 to December 2015. The original data and the scripts required to join them into the final dataset is available online from the authors at https://github.com/jlospinoso/uvi-sunburn.

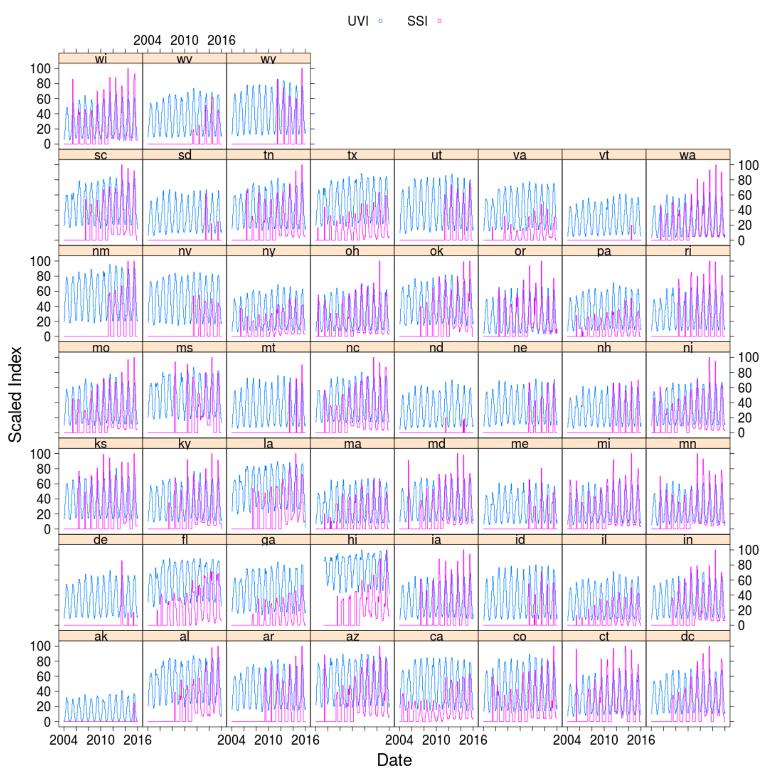

Figure 1. *Ultraviolet Index (UVI) and Google Sunburn Search Index (SSI) time series by state. Time series is given at monthly resolution. UVI (in pink) and SSI (in blue) are scaled to fit onto 0 to 100 y-axis.*

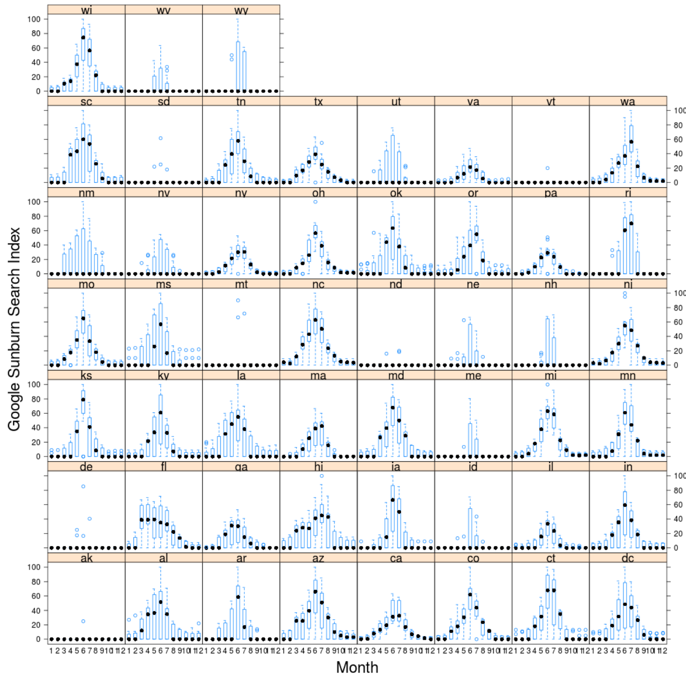

Figure 2. *Google Sunburn Search Index by state. Box and whisker plots are given by month. The box represents the first through third inter-quartile range, and dots represent the mean value for the given month.*

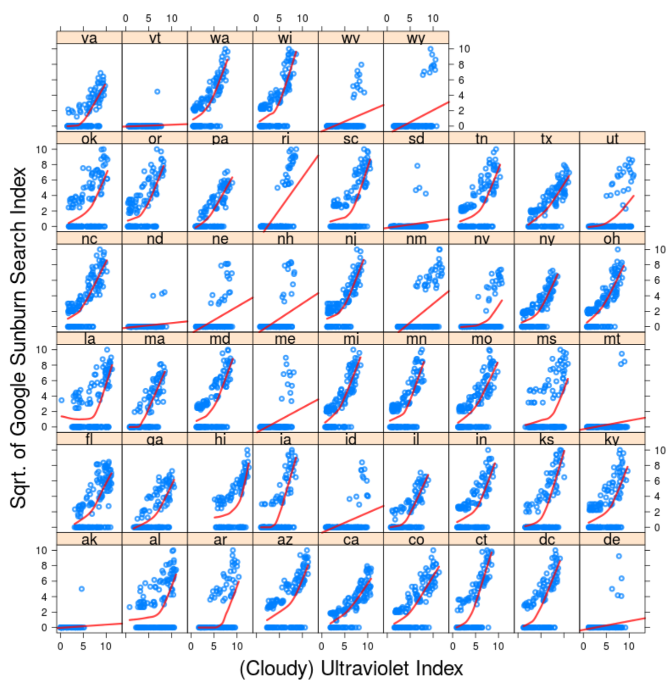

Figure 3. *Google SSI vs. UVI with variance stabilizing transformation applied [36]. LOESS regression or linear regression given in red [37].*

The UVI and SSI time series are given in **Figure 1**. As expected, the UVI series is strongly periodic and has a seasonal nature. In general, the series do not appear to drift much from one year to the next and they are highly regular. There is clearly variation between states. Compare, e.g., Alaska (AK) and Arizona (AZ); the amplitude, or peak UVI, of Arizona is far greater, as intuition might predict. Compared with the UVI time series, there is far more heterogeneity between states for SSI. For a number of states, notably Alaska (AK), Delaware (DE), North Dakota (ND), Vermont (VA), and South Dakota (SD) there is barely any non-zero SSI in the series. For others like Arizona (AZ), Idaho (ID), Maine (ME), New Hampshire (NH), Nebraska

(NE), Montana (MT), Nevada (NV), New Mexico (NM), Utah (UT), West Virginia (WV), and Wyoming (WY), the SSI is zero until roughly half-way through the observational period. One possible explanation for this non-stationarity in the time series is the adoption of internet searching into these areas. Regardless, it will be important during analysis to allow for this heterogeneity between (and within) states. It is fairly clear from **Figure 1** that there is a strong, seasonal correlation between SSI and UVI.

Despite substantial difference among states in the magnitude of SSI across time series, the seasonality is still fairly discernible. **Figure 2** gives box and whisker plots for each month's SSI, given by state. It is clear that SSI is elevated during the months of May to July virtually across the board. For some other states, e.g. Alabama, Hawaii, Florida, New Jersey, and South Carolina, the SSI increases even earlier during the months of March and April. One possibility is that these states are coastal and represent vacation destinations for beach goers, who are at particular risk for sunburn.

An important question is whether or not Google's SSI is driven by the UVI. **Figure 3** illustrates SSI vs. UVI at the state level. The hypothesis is that there is a positive relationship between these two variables (i.e. a positive slope). **Figure 3** contains both the raw data (as blue, circular points) and a regression line (as red, solid lines) for each state. As expected, there is significant heterogeneity in this relationship at the state level. It will be crucial to account for this heterogeneity in the statistical modeling process. Nonetheless, there appears to be a very strong, positive relationship between UVI and SSI for most states.

The collected Google SSI forms so-called panel data [29] where multiple time series are present. Since the data contains time series, it is required to control for autocorrelation, i.e. the dependency of observed search activity on previous search activity. Failure to account for autocorrelation present in the data can cause serious inferential problems [30-33]. Accordingly, statistical modeling of SSI requires careful consideration. We are interested in how the UVI index drives the frequency of search terms; if we fail to specify a reasonable statistical model, we would risk obtaining erroneous results.

Since each state has a time series, it may be required to incorporate temporal features such as autoregressive, integrative, and moving-average (ARIMA) coefficients [29, 34]. It is possible to diagnose autocorrelation issues through analysis of the partial autocorrelation function. If an inadequate model specification is found, modifications can be made. This three-stage modeling approach follows the popular Box-Jenkins methodology [35].

The form of the linear model entails capturing the dependency of SSI on the covariates through its mean:

mean(Sit)=α+αi+αt+(β+βi)Uit+γ1Sit-1+γ2Sit-2+...

in which Uit is the UVI, β is the UVI coefficient, βi is its state-specific counterpart, α is constant, α is a state-specific constant, αt is a time-specific constant, and γ are so-called autocorrelation coefficients. The variable of interest is β, or UVI coefficient. If strongly positive, it would indicate that SSI increases when the UVI increases. As is generally advised in the statistical literature, variance-stabilizing transformations [36] of the covariates will be tested for improved fit. Such transformations can have a dramatic effect on residuals, possibly mitigating obvious violations of model assumptions, e.g. that the residuals are normally distributed [37].

The final model is presented in **Table 1**. A square root transformation for all UVI- and SSI-related terms yielded superior model fit. A monthly lag for both UVI and SSI, coupled with an 11- and 12-month lag term for SSI removed all partial autocorrelations out to a year. Random effects [38] were grouped at the state level and included an intercept, the UVI coefficient, and the monthly UVI lag coefficient. Fixed effects [39] were included for each month (January is the base case) and each year (2005 is the base case).

The contemporaneous UVI had a strong, positive, and statistically significant effect on the SSI. For each one-unit increase in the square root of UVI, a 2.30-unit increase in the square root of SSI is expected. The UVI lag effect is negative, but smaller than the contemporaneous effect (-1.58). This suggests

Table 1. *Estimation results from a mixed-effects maximum likelihood estimation on SSI.*

	Estimate	Std. Err.	p-value
UVI	2.30	0.19	0.00
UVI lag	-1.58	0.18	0.00
SSI lag 1	0.27	0.01	0.00
SSI lag 11	0.06	0.01	0.00
SSI lag 12	0.20	0.01	0.00
Intercept	-1.68	0.22	0.00
February	-0.45	0.10	0.00
March	0.02	0.13	0.87
April	0.48	0.17	0.00
May	1.39	0.21	0.00
June	2.52	0.24	0.00
July	1.37	0.25	0.00
August	-0.06	0.25	0.81
September	0.02	0.22	0.94
October	0.52	0.18	0.00
November	0.79	0.13	0.00
December	0.68	0.10	0.00
2006	-0.16	0.12	0.18
2007	0.15	0.12	0.21
2008	0.28	0.11	0.02
2009	0.31	0.12	0.01
2010	0.50	0.12	0.00
2011	1.18	0.12	0.00
2012	1.08	0.12	0.00
2013	1.09	0.12	0.00
2014	1.03	0.13	0.00
2015	1.02	0.12	0.00
2016	1.10	0.21	0.00

would-be sunburn searchers tend to search even more when there is a sharp increase in UVI when compared to a consistently high UVI. There is a moderate autocorrelation in SSI; for every one-unit increase in the square root of SSI, a 0.26-unit increase is expected in the square root of the following month's SSI. Interestingly, there is an annual component to the SSI's autocorrelation, as we expect an additional 0.20 unit increase a year later.

There are yearly fixed effects. From 2007 to 2011, there is a steady increase in SSI in general. From 2012 onward, the fixed effects seem to be roughly equal. There are also monthly fixed effects. During the months of April and May, SSI is elevated. The monthly fixed effect peaks during June (+2.52 over January) and backs off a bit in July (+1.37). That August and September do not have a statistically significant difference from January lends additional evidence to the argument that increases in UVI have compounding effects on SSI.

Discussion

As expected, UVI has a strong, positive, and statistically significant effect on SSI. The nature of the lag terms indicates that sharp increases in UVI tend to increase SSI more than sustained, high levels of UVI. One possible interpretation of this evidence is the following: when UVI rises sharply, the prevalence of sunburn also increases sharply. Once an individual has developed a sunburn, he/she will tend to take steps to mitigate sunburn in the immediate future. The heavily seasonal nature of the fixed-effects lends some additional evidence to this explanation; during April, May, and June—months that are associated with activities where sun exposure could be elevated—there are additional contributions to SSI (even controlling for UVI and lag effects). Essentially, the early months of the sunning season could tend to catch people off-guard. When compared to July and August—also months associated with elevated sun exposure—this effect disappears, lending evidence to the notion that once a person has been sunburned, he/she is less likely to get burned again later in the season.

Conclusion

Having established that UVI does indeed drive SSI, there is strong evidence that SSI can be used as a proxy for the prevalence of sunburn. Accordingly, it would be instructive to refer to the nature of the searches about sun exposure. Future work could analyze the interplay between searches for e.g. "sunburn" and "sunblock." Since both SSI and UVI can be obtained at fairly granular geographic levels, it may be possible to use them as instruments to measure the effectiveness of sunburn awareness campaigns. For example, if searches for sunburn prophylaxis increase whereas searches about sunburn remedies decrease after an intervention, it could be used as evidence to substantiate the effectiveness of the intervention.

References

1. Turgay AS, Sari D, Can M, Genc RE. Determination of sunburn and skin cancer risk of outpatients in a dermatology polyclinic. Asian Pacific J Cancer Prev. 2005;6(2):143-146. [PMID:16101323].
2. Achachi A, Vocanson M, Bastien P, et al. UV Radiation Induces the Epidermal Recruitment of Dendritic Cells that Compensate for the Depletion of Langerhans Cells in Human Skin. J Invest Dermatol. 2015;135(8):2058-2067. doi:10.1038/jid.2015.118. [PMID:25806853].
3. Murphy G, Young a R, Wulf HC, Kulms D, Schwarz T. The molecular determinants of sunburn cell formation. Exp Dermatol. 2001;10(3):155-160. [PMID:11380610].
4. Gilchrest BA, Soter NA, Stoff JS, Mihm MC. The human sunburn reaction: Histologic and biochemical studies. J Am Acad Dermatol. 1981;5(4):411-422. [PMID:7287956].
5. Saraiya M, Hall HI, Uhler RJ, et al. Sunburn prevalence among adults in the United States, 1999. Am J Prev Med. 2002;23(2):91-97. doi:10.1016/S0749-3797(02)00461-0. [PMID:12121796].
6. Saraiya M, Glanz K, Briss PA, et al. Interventions to prevent skin cancer by reducing exposure to ultraviolet radiation: A systematic review. Am J Prev Med. 2004;27(5):422-466. [PMID:15556744].
7. Buller DB, Cokkinides V, Hall HI, et al. Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: Review from national surveys and case studies of 3 states. J Am Acad Dermatol. 2011;65(5 SUPPL. 1):S114.e1-S114.e11. [PMID:22018060].
8. Varada S, Alessa D. Saving Their Skins. JAMA Dermatology. 2014;150(9):989. [PMID:25207531].
9. Hiemstra M, Glanz K, Nehl E. Changes in sunburn and tanning attitudes among lifeguards over a summer season. J Am Acad Dermatol. 2012;66(3):430-437. [PMID:3728666].
10. Castle CM, Skinner TC, Hampson SE. Young women and suntanning: An evaluation of a health education leaflet. Psychol Health. 1999;14(3):517-527. doi:10.1080/08870449908407344.
11. Dobbinson SJ, Volkov A, Wakefield MA. Continued Impact of SunSmart Advertising on Youth and Adults' Behaviors. Am J Prev Med. 2015;49(1):20-28. [PMID:25794474].
12. Dobbinson SJ, Wakefield MA, Jamsen KM, et al. Weekend Sun Protection and Sunburn in Australia. Am J Prev Med. 2008;34(2):94-101. [PMID:18201638].
13. Boldeman C, Bränström R, Dal H, et al. Tanning habits and sunburn in a Swedish population age 13-50 years. Eur J Cancer. 2001;37(18):2441-2448. [PMID:11720841].
14. Stott MA. Tanning and sunburn: Knowledge, attitudes and behaviour of people in Great Britain. J Public Health Med. 1999;21(4):377-384. [PMID:11469358].
15. Robinson JK, Rademaker AW. Sun Protection by Families at the

Beach. Arch Pediatr Adolesc Med. 1998;152(5). [PMID:9605030].
16. Craig S. Long and Alvin J. Miller, Climate Prediction Center, Hai-Tien Lee and Jeannette D. Wild, Research and Data Systems I. UltraViolet Index. The National Weather Service. http://www.nws.noaa.gov/om/tpb/uvindex.htm. Published 1994. Accessed April 1, 2016.
17. Carneiro HA, Mylonakis E. Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. Clin Infect Dis. 2009;49(10):1557-1564. [PMID:19845471].
18. Willson TJ, Shams A, Lospinoso J, Weitzel E, McMains K. Searching for Cedar: Geographic Variation in Single Aeroallergen Shows Dose Response in Internet Search Activity. Otolaryngol -- Head Neck Surg. 2015;153(5):770-774. [PMID:26340925].
19. Willson TJ, Lospinoso J, Weitzel EK, McMains KC. Effect of environmental factors on Internet searches related to sinusitis. Laryngoscope. 2015;125(11):2447-2450. [PMID:26340925].
20. Willson TJ, Lospinoso J, Weitzel E, McMains K. Correlating Regional Aeroallergen Effects on Internet Search Activity. Otolaryngol -- Head Neck Surg. 2015;152(2):228-232. [PMID:25505261].
21. Braun T, Harréus U. Medical nowcasting using Google trends: application in otolaryngology. Eur Arch Oto-Rhino-Laryngology. 2013;270(7):2157-2160. [PMID:23632877].
22. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for Influenza Surveillance in South China. PLoS One. 2013;8(1):2009-2014. [PMID:23372837].
23. Ginsberg J, Mohebbi MH, PaTel:RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012-1014. [PMID:19020500].
24. Kaidbey KH, Grove KH, Kligman AM. The influence of longwave ultraviolet radiation on sunburn cell production by UVB. J Invest Dermatol. 1979;73(3):743-745. [PMID:469280].
25. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988;124(6):869-871. [PMID:3377516].
26. Thieden E, Agren MS, Wulf HC. Solar UVR exposures of indoor workers in a Working and a Holiday Period assessed by personal dosimeters and sun exposure diaries. Photodermatol Photoimmunol Photomed. 2001;17(6):249-255. [PMID:11722749].
27. Westerdahl J, Olsson H, Ingvar C. At what age do sunburn episodes play a crucial role for the development of malignant melanoma. Eur J Cancer. 1994;30(11):1647-1654. [PMID:7833138].
28. Brash DE, Ziegler A, Jonason AS, Simon JA, Kunala S, Leffell DJ. Sunlight and sunburn in human skin cancer: p53, apoptosis, and tumor promotion. J Investig Dermatol Symp Proc. 1996;1(2):136-142. [PMID:9627707].
29. Wooldridge J. Econometric Analysis of Cross Section and Panel Data. MIT Press; 2002.
30. Bence JR. Analysis of short time series: Correcting for autocorrelation. Ecology. 1995;76(2):628-639. doi:10.2307/1941218.
31. King ML, Giles DEA. Autocorrelation pre-testing in the linear model. J Econom. 1984;25(1-2):35-48. doi:10.1016/0304-4076(84)90035-6.
32. Hogenraad R, McKenzie DP, Martindale C. The enemy within: Autocorrelation bias in content analysis of narratives. Comput Hum. 1997;30(6):433-439. doi:10.1007/BF00057939.
33. Park RE, Mitchell BM. Estimating the autocorrelated error model with trended data. J Econom. 1980;13(2):185-201. doi:10.1016/0304-4076(80)90014-7.
34. Brockwell PJ, Davis RA. Time Series: Theory and Methods. (3rd Edition), Springer, 1991.
35. Box G, Jenkins GM, Reinsel G. Time Series Analysis: Forecasting & Control (3rd Edition). Holden-Day, 1994.
36. Box GEP, Cox DR. An analysis of transformations revisited, rebutted. J R Stat Soc Ser B-Statistical Methodol. 1982;77(377):209. http://www.jstor.org/stable/2984418. Accessed April 12, 2016.
37. Cleveland WS, Devlin SJ. Locally Weighted Regression : An Approach to Regression Analysis by Local Fitting Locally Weighted Regression : An Approach to Regression Analysis by Local Fitting. J Am Stat Assoc. 1988;83:403, 596-610. doi:10.1080/01621459.1988.10478639.
38. Lazega E, Snijders T. Multilevel Network Analysis for the Social Sciences. (Lazega E, Snijders TAB, eds.). Cham: Springer International Publishing; 2016. doi:10.1007/978-3-319-24520-1.
39. Wooldridge J. Introductory Econometrics. 3E ed. McGraw Hill, 2005.