

Bootstrap Approach to Correlation Analysis of Two Mineral Components

T.O. Olatayo

Department of Mathematical Science, Faculty of Science,
Olabisi Onabanjo University, Ago-Iwoye P.M.B.2002, Ogun State, Nigeria

Abstract: Problem statement: In this article we considered pairs bootstrap through a truncated geometric bootstrap method for stationary time series data. Construction of valid inferential procedures through the estimates of standard error, coefficient of variation and other measures of statistical precision such as bootstrap confidence interval were considered. The method was used to confirm the correlation between Silicon Oxide (SiO_2) and Aluminum Oxide (Al_2O_3) from a geological data. A typical problem is that can these components exist together or they are mutually exclusive. **Approach:** We attempt to solve these problems through bootstrap approach to correlation analysis and show that pair bootstrap method through truncated geometric bootstrap method for stationary process revealed the correlation coefficient between Silicon Oxide (SiO_2) and Aluminum Oxide (Al_2O_3) from the same geological field. **Results:** The computed measure of statistical precisions such as standard error, coefficient of variation and bootstrap-t confidence interval revealed the correlation analysis of the bivariate stochastic processes of SiO_2 and Al_2O_3 components from the same geological field. **Conclusion:** The correlation analysis of the bivariate stochastic process of SiO_2 and Al_2O_3 components through bootstrap method discussed in this study revealed that the correlation coefficients are negative and bootstrap confidence intervals are negatively skewed for all bootstrap replicates. This implies that as one component increases, the other component decreases, which means that the two components are mutually exclusive and the abundance of one mineral prevents the other in the same oil reservoir of the same geological field.

Key words: Pair bootstrap, standard error, coefficient of variation, bootstrap-t confidence interval, correlation analysis, Silicon Oxide (SiO_2), Aluminum Oxide (Al_2O_3)

INTRODUCTION

Since its introduction by Efron (1979), the bootstrap has become a method of choice for assessing uncertainty in a vast range domain. One of the most often used statistical tools, not only in geosciences, is Pearson's correlation coefficient r_{xy} , which measures the degree of (linear) interrelation between two sample (data size n) variables, x and y . Quite frequently x and y are measured over time and a typical aim in correlation analysis of such bivariate time series is to value the evidence for an influence of one time-dependent variable on the other. Mudelsee (2003) analyzed the influence of solar activity on monsoon rainfall during early Holocene. Since geological interpretation of a detected correlation requires knowledge about the statistical precision, a confidence interval for r_{xy} or at least, a test of the hypothesis "population correlation coefficient $r_{xy} = 0$ " is required.

For a typical geological or climatologically time series, estimation of a confidence interval for r_{xy} is hindered by positive serial dependence; Wilks (1995)

and Storch and Zwiers (2001). The stationary bootstrap of Politis and Romano (1994) and the moving block bootstrap of Kunsch (1989) are obvious methods to solve these problems.

The present study employs the use of pairs bootstrap through truncated geometric bootstrap method of Olatayo (2010), with block length proportional to the estimated time dependent of the data. The study is based on the use of some chemical components of the soil to determine the presence or the abundance of oil in an area under exploration. The field extraction is done by digging and a sample of sand is collected and experiments carried out to determine the components in it, their quantity and the type of soil it is during chemostratigraphy study. These components are the principal oxide of sand or sand stone, which is SiO_2 or silicon oxide and the principle oxide of clay or sill stone which is Al_2O_3 or Aluminum oxide. The point is that the bulk of oil reservoir rocks in Nigeria sedimentary basins are sandstone and shale a product of sill stone or mud stone Olanrewaju (2007) and Nwachukwu (2007). The geological data from the core

of about 6,000 m is from chemostratigraphic data from Batan well at 30 m regular interval, Sofowora (2002). A typical problem is that, can these components exist together or they are mutually exclusive. We attempt to solve the problem through bootstrap approach to correlation analysis.

MATERIALS AND METHODS

Suppose $\{X_t, Y_t\}$ are two interacting time series. Then we can regard this pair of time series as realization of a hypothetical population of pairs of time series, called a bivariate stochastic process $\{X_t, Y_t\}$. We shall assume that the data are read off at equispaced times yielding a pair of discrete time series, generated by a discrete bivariate process and that values of the time series at times $t_{0+h}, t_{0+2h}, \dots, t_{0+Nh}$ are denoted by $(X_t, Y_t), (X_2, Y_2), \dots, (X_N, Y_N)$. In general, a bivariate stochastic process $\{X_t, Y_t\}$ need not be stationary, however we assume that the appropriately differenced process $\{x_t, y_t\}$ is stationary. If in addition, it is assumed that a bivariate process is Gaussian or Normal, then it is uniquely characterized matrix, Box and Jenkins (1976). Therefore we can analyze by viewing them as being causally related.

The cross covariance function between X_t and Y_t at lag k is Eq. 1:

$$\gamma_{xy}(k) = E[x_t - \mu_x][y_{t+k} - \mu_y] \quad k = 0, 1, 2, \dots \quad (1)$$

The function $\gamma_{xy}(k)$ is called the cross covariance function of the bivariate process. Similarly Eq. 2:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\gamma_x \gamma_y} \quad k = 0, \pm 1, \pm 2, \dots \quad (2)$$

Is called the correlation coefficient at lag k and the function $\rho_{xy}(k)$, is the cross correlation function of the bivariate process.

Pairs bootstrap: Resampling from the matrix with typical row $\{X_t, Y_t\}$ whereby we will no longer condition on the X_t since each bootstrap sample now has a different value. First, we randomly select n pairs of samples with replacement from the original observation $\{x_1^1 y_1^1\}, \dots, \{x_1^1 y_1^1\}$, where $t = 1 \dots n$. The resulting bootstrap sample is $\{x_1^* y_1^*\}, \dots, \{x_t^* y_t^*\}$. The pair bootstrap is valid when the errors display

heteroscedasticity of unknown form; it works even for dynamic models, Freedman (1984).

The vicariate stochastic data where x_t is the SiO_2 silicon oxide while y_t is the Al_2O_3 Aluminum oxide. Using pair bootstrap with truncated geometric distributed block length of Olatayo (2010), the consecutive pairs are generated as follows:

Let $\{x_{(i)}^*, y_{(i)}^*\} = \{x_{(j)}, y_{(j)}\}$ with probability $(1-p)$, with p small, the successor of $\{x_{(i)}^*, y_{(i)}^*\}$ is taken to be the Successor of $(x_{(j+1)}, y_{(j+1)})$ and with probability P , $(x_{(j+1)}, y_{(j+1)})$ is a randomly selected pair of the original time series, that is, $(x_{(i+1)}^*, y_{(i+1)}^*) = [x_{(j+1)}^*, y_{(j+1)}^*]$ and with probability P , $(x_{(i+1)}^*, y_{(i+1)}^*)$ is a randomly selected pair of the original vicariate time series. Next, from the resample data the bootstrap replicate, γ_{xy}^* is calculated which is repeated until B replicates exist. That is, independent repetitions of the bootstrap sampling process give bootstrap replications $\gamma_{xy}^*(1), \gamma_{xy}^*(2), \dots, \gamma_{xy}^*(B)$ respectively. We then compute the correlation coefficient, standard error, coefficient of variation and bootstrap-t confidence interval as a measure of statistical precision Olatayo *et al.* (2010).

RESULTS

The summary of our findings on the performance of a pair bootstrap through the truncated geometric bootstrap method as described in section 2 and bootstrap replicates of ($B = 50, 100$ and 250) with measures of statistical precision of correlation coefficient from a geological data is presented.

The Table 1 reveals the standard error S_{eB} for correlation coefficient between the two bivariate variables. The correlation coefficient and Coefficient of Variation (C.V.) at each bootstrap replicates are negative meaning that as one mineral that is SiO_2 is increasing the other one Al_2O_3 is decreasing.

The Confidence Interval (CI), following the prescribed bootstrap t-confidence interval of Olatayo *et al.* (2010) is presented in Table 2 above. It revealed that the distribution of the correlation coefficient between the two variables is negatively skewed.

Table 1: Summary statistics for bootstrap correlation coefficient

B	ρ_{xy}	S_{eB}	C.V.
50	-0.5213	0.0614	-0.1178
100	-0.5181	0.0688	-0.1328
250	-0.4642	0.0725	-0.1562

Table 2: Summary Statistics of 95% bootstrap confidence interval for bootstrap correlation coefficient

B	ρ_{xy}	Se _B	95% C.I.
50	-0.5213	0.0614	[-0.6447, -0.3980]
100	-0.5181	0.0688	[-0.6546, -0.3816]
250	-0.4642	0.0725	[-0.6063, -0.3199]

CONCLUSION

The correlation analysis of the vicariate stochastic process of SiO₂ and Al₂O₃ components through bootstrap method discussed in this study revealed that the correlation coefficient is negative for all bootstrap B replicates which implies that as one variable increases the other decrease, that is, the two components SiO₂ and Al₂O₃ are mutually exclusive and the abundance of one mineral prevents the other in the same oil reservoir of the same geological field.

REFERENCES

Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco, ISBN-10: 0816211043, pp: 575.

Efron, B., 1979. Bootstrap methods: Another look at the Jackknife. *Ann. Statist.*, 7: 1-26.

Freedman, D., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.*, 12: 827-842.

Kunsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17: 1217-1241.

Mudelsee, M., 2003. Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Math. Geol.*, 35: 651-665. DOI: 10.1023/B:MATG.0000002982.52104.02

Nwachukwu, J.I., 2007. Organic Matter: The Source of our Wealth. 1st Edn., Obafemi Awolowo University Press, Ile-Ife, Nigeria, pp: 60.

Olanrewaju, V.O., 2007. Rocks: Their beauty, language and roles as resources of economic development. Obafemi Awolowo University, Ile-Ife, Nigeria.

Olatayo, T.O., 2010. On truncated geometric bootstrapping method for stochastic time series process. An Unpublished Ph.D. Thesis, Department of Statistics, University of Ibadan, Ibadan, Nigeria.

Olatayo, T.O., G.N. Amatia and T.O. Obilade, 2010. Bootstrap method for Dependent data structure and measure of statistical precision. *J. Math. Statist.*, 6: 84-88. DOI: 10.3844/jmssp.2010.84.88

Politis, D.N. and J.P. Romano, 1994. The stationary bootstrap. *J. Am. Statist. Assoc.*, 89: 1303-1313. DOI: 10.1080/01621459.1994.10476870

Sofowora, K.T., 2002. Factor Analysis: Application with a chemostratigraphic data. Obafemi Awolowo University, Ile-Ife, Nigeria.

Storch, H.V. and F.W. Zwiers, 2001. Statistical Analysis in Climate Research. 1st Edn., Cambridge University Press, Cambridge, UK., ISBN-10: 0521012309, pp: 484.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences: An Introduction. 1st Edn., Academic Press, San Diego, CA., ISBN-10: 0080541720, pp: 467.