

Bootstrap Approach to Correlation Analysis of Two Mineral Components in a Geological Field

*T.O. OLATAYO, **A.A. AKOMOLAFE and *A.I. TAIWO

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

otimtoy@yahoo.com

**Department of Mathematics and Statistics, Joseph Ayo Babalola University
Ikeji-Arakeji, Osun State, Nigeria.

akomolafea@yahoo.com

Abstract

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 Aluminium Oxide (Al_2O_3) from a geological data. A typical problem is that can these components exist together or they are mutually exclusive. We attempt to solve these problem 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 Aluminium oxide (Al_2O_3) from the same geological field. 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.

The correlation analysis of the bivariate stochastic process of SiO_2 and Al_2O_3 components through bootstrap method discussed in this paper revealed that the correlation coefficient 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 mean 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 and correlation analysis.

1. INTRODUCTION

Since its introduction by Efron (1979), the bootstrap has become a method of choice for assessing uncertainty in a vast range domain, Davison et al., (2003). One of the most often used statistical tools, not only in geosciences, is person's correlation coefficient Γ_{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) analysed 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 Γ_{xy} or at least, a test of the hypothesis "population correlation coefficient $\rho_{xy} = 0$ " is required.

For a typical geological or climatological time series, estimation of a confidence interval for Γ_{xy} is hindered by positive serial dependence; Wilks (1995), Von and Zwiers (1999). 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 paper 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 Aluminium oxide. The point is that the bulk of oil reservoir rocks in Nigeria sedimentary basins is sandstone and shale a product of sill stone or mud stone Olanrewaju (2007) and Nwachukwu (2007). The geological data from the core of about 6,000m is from chemostratigraphic data from Batan well at 30m 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 pair bootstrap approach to correlation analysis.

2. 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 (1970). Therefore we can analyse by viewing them as being causally related.

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

$$\partial_{xy}(k) = E[x_t - \mu_x][y_{t+k} - \mu_y] \dots\dots\dots (2.1)$$

$$k=0, 1, 2, \dots$$

The function $\partial_{xy}(k)$ is called the cross covariance function of the bivariate process.

Similarly,

$$\rho_{xy}(k) = \frac{\partial_{xy}(k)}{\partial_x \partial_y} \quad k = 0, \pm 1, \pm 2, \dots \quad \dots \dots \dots (2.2)$$

Is called the cross 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 $\{x_1^1 y_1^1\}, \dots, \{x_t^1 y_t^1\}$, 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 bivariate stochastic data where x_t is the SiO₂ silicon oxide while y_t is the Al₂O₃ Aluminium 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 bivariate time series. Next, from the resampled data the bootstrap replicate, $\hat{\rho}_{xy}^*$ is calculated which is repeated until B replicates exist. That is, independent repetitions of the bootstrap sampling process give bootstrap replications $\hat{\rho}_{xy}^*(1), \hat{\rho}_{xy}^*(2), \dots, \hat{\rho}_{xy}^*(B)$ respectively. We then compute the standard error, coefficient of variation and bootstrap-t confidence interval as a measure of statistical precision Olatayo et al., (2010).

3.Results and Discussions

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.

Table 1: Summary Statistics for Bootstrap Correlation Coefficient.

B	Mean	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

The above table 1 reveal 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.

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

B	Mean	S_{eB}	95% C.I.
50	-0.52123	0.0614	[-0.6447, -0.3980]
100	-0.5181	0.0688	[-0.6546, -0.3816]
250	-0.4642	0.0725	[-0.6063, -0.3199]

The confidence interval (C.I.), 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 are negatively skewed.

4. Conclusion

The correlation analysis of the bivariate stochastic process of SiO_2 and Al_2O_3 components through bootstrap method discussed in this paper 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 SiO_2 and Al_2O_3 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. and Jenkins G.M. (1970). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
 Efron, B. (1979). Bootstrap Methods another look at eh Jacknife. Ann. Statist. 7, 1-26.

- Freedman, D.A. (1984) On Bootstrapping Two-stage least squares estimates in stationary linear models. *Annals of Statistics* 12, 827-842.
- Kunsch, H. (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: *Mathematical Geology*, Vol. 35, No. 6, 651-665.
- Nwachukwu, J.I. (2007) Organic Matter, the source of our wealth. An inaugural lecture delivered at Oduduwa Hall, Obafemi Awolowo University, Ile-Ife, Nigeria.
- Olanrewaju, V.O. (2007) Rocks: Their beauty, language and roles as resources of economic development. An inaugural lecture delivered at Oduduwa Hall, 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., Amatia, G.N. and Obilade, T.O. (2010) Bootstrap method for Dependent data structure and measure of statistical precision. *Journal of Mathematical and Statistics* 6(2) 84-88.
- Politis, D.N. and Romano, J.O. (1994) The Stationary Bootstrap. *Journal of American statistical Association* 89, 1303-1313.
- Sofowora, K.T. (2002) Factor Analysis: Application with a chemostratigraphic data. A project submitted to the Department of Mathematics, Faculty of Science, Obafemi Awolowo University, Ile-Ife, Nigeria.
- Von Storch, H. and Zwiers, F.W. (1999) *Statistical Analysis in climate research*: Cambridge University Press, Cambridge, UK, 488p.
- Wilks, D.S. (1995) *Statistical Methods in the atmospheric science*; Academic Press, San Diego, CA, 467 p.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage:

<http://www.iiste.org>

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** <http://www.iiste.org/Journals/>

The IISTE editorial team promises to review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

