Journal of Multivariate Analysis 144 (2016) 81–90

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure

Anuradha Roy [a,*], Roman Zmyślony [b], Miguel Fonseca [c], Ricardo Leiva [d]

[a] Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX 78249, USA

[b] Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Szafrana 4a, 65-516 Zielona Góra, Poland

[c] Centro de Matemática e Aplicações, Universidade Nova de Lisboa, Monte da Caparica, 2829-516 Caparica, Portugal

[d] Departamento de Matemática, F.C.E., Universidad Nacional de Cuyo, 5500 Mendoza, Argentina

ARTICLE INFO

Article history:
Received 23 February 2015
Available online 1 December 2015

AMS 2010 subject classifications:
62H12
62J10
62F10

Keywords:
Best unbiased estimator
Blocked compound symmetric covariance structure
Doubly multivariate data
Coordinate free approach

ABSTRACT

The paper deals with the best unbiased estimators of the blocked compound symmetric covariance structure for m-variate observations over u sites under the assumption of multivariate normality. The free-coordinate approach is used to prove that the quadratic estimation of covariance parameters is equivalent to linear estimation with a properly defined inner product in the space of symmetric matrices. Complete statistics are then derived to prove that the estimators are best unbiased. Finally, strong consistency is proven. The proposed method is implemented with a real data set.

1. Introduction

This article deals with the estimation and best unbiased estimators of the blocked compound symmetric (BCS) covariance structure or block exchangeable covariance structure (defined in Section 2) for doubly multivariate observations (m dimensional observation vector repeatedly measured over u locations or time points). For doubly multivariate observations number of repeated multivariate observations u must be greater than 1; for $u = 1$, the data just become multivariate data. Doubly multivariate data require extraction of relevant information that is hidden in the data in order to model the data appropriately and accurately. Blocked compound symmetric covariance structure for doubly multivariate observations is a multivariate generalization of compound symmetric covariance structure for multivariate observations. It was introduced by Rao [22,23] while classifying genetically different groups. Arnold [1] considered testing problems in BCS covariance setting. He proposed the orthogonalization of the problem, and suggested testing by a product of independent beta-variates. Another major contribution was made by Arnold [2], who studied this BCS covariance structure while developing general linear model with exchangeable and jointly normally distributed error vectors. Overview of all previous results was given in [29]. Rogers and Young [25] derived maximum likelihood estimators of unknown parameters via orthogonalization, but only for the component matrices $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ in BCS structure having compound symmetry structure. They also proposed tests for some variants of the covariance structure. Multivariate measures of interclass and intraclass correlations for more

* Corresponding author.
E-mail address: Anuradha.Roy@utsa.edu (A. Roy).

http://dx.doi.org/10.1016/j.jmva.2015.11.001

than one characteristic (multivariate observations) are introduced in [11,12] to assess the degree of resemblance between family members. These authors also discussed the inferential procedures for interclass and intraclass correlations in the multivariate situation for more than one characteristic by proposing some unified estimators and derived the asymptotic distributions of these estimators. Afterwards BCS covariance structure did not attract much attention in the literature for some time until Leiva [15] developed classification rules for doubly multivariate observations and generalized Fisher's linear discrimination method assuming BCS covariance structure, which he named as equicorrelated covariance structure, for the data. Leiva [15] derived maximum likelihood estimates (MLEs) of the BCS covariance structure and developed classification rules using these MLEs. Lately, this covariance structure started gaining a lot of attention in the literature, especially in the area of high-dimensional estimation (see [27]). Leiva and Roy [16] derived a quadratic extension of the traditional classification rules for multiple classes based on equicorrelated training vectors by using BCS covariance structure. Liang et al. [17] developed mixed linear model with BCS covariance structure in conjunction with a symmetric circular Toeplitz matrix in each block as random error, and derived explicit MLEs of the unknown parameters in this setting. Recently, Roy et al. [28] obtained a natural extension of the Hotelling's T^2 statistic, the Block T^2 statistic, a convolution of two Hotelling's T^2's, using unbiased estimators of the component matrices of the orthogonally transformed BCS covariance matrix while testing the equality of mean vectors for paired doubly multivariate observations. Hao et al. [8] obtained the eigendecomposition of the BCS covariance structure while finding the principal components of multivariate interval data. They assumed all vertices of the hyper-rectangle defined by the intervals of all variables for each observation has BCS structure, and then derived unbiased estimators of the eigenblocks of the BCS structure. More recently, Žežula et al. [32] used unbiased estimators of the component matrices of the orthogonally transformed BCS covariance matrix for testing of equality of mean vectors in two independent populations using Block T^2 statistic for multivariate repeated measures data. To the best knowledge of the authors, none of the previous studies have considered the estimation properties of the BCS covariance structure. A natural question then is whether or not these estimators are "good" in some sense.

One measure of "good" is "unbiasedness". This article derives the unbiased estimators for parameters of the unstructured mean vector and the BCS covariance structure following the same way as Roy et al. [28], and addresses the issue of optimal properties of these unbiased estimators that is motivated by real-world applications. A characterization of best linear unbiased estimator (BLUE) given by Zmyślony [34] and completeness in [35] are used to derive the optimal properties of unbiased estimators of BCS covariance matrix. The derivation and computation of these estimators are developed using the coordinate free approach (see [13,3]).

An important advantage of using BCS structure for doubly multivariate data is that the number of unknown parameters is only $m(m+1)$, which does not even depend on the number of repeated measures u, whereas the number of unknown parameters in the $(um \times um)$-dimensional unstructured covariance matrix is $um(um+1)/2$, which can increase very rapidly with the increase of either m or u. Hence, BCS covariance structure allows the number of repeated measurements u to grow unrestrictedly, and thereby provides more information, while the number of unknown parameters remains the same.

The BCS structure of covariance matrix, actually, is a common structure in real data analysis; for example in analyzing longitudinal data, multivariate spatial data, and multivariate time series data. Consider an example of a longitudinal data where researchers measure levels of fat byproducts repeatedly over time, say for eight consecutive weeks, in a clinical trial study of some skin care product. Consider another example from multivariate spatial data of a digital image where contiguous pixels are correlated. A pixel is generally thought of as the smallest single component of a digital image. The intensity of each pixel is variable; in color systems, each pixel has typically three or four components such as red, green, and blue where each pixel is a point in RGB space. Thus, RGB space is a three-dimensional vector space, and each pixel is defined by an ordered triple of red, green, and blue coordinates (m). Each primary color component (red, green, or blue) represents an intensity which varies linearly from 0 to a maximum value, which corresponds to full saturation of that color. Correlations exist among the intensities of the ordered triple of red, green, and blue coordinates. As mentioned before, correlation also exists among the contiguous pixels (u), because sensors take a significant energy from these contiguous pixels and sensors cover a land region much larger than the size of a pixel. For example, if a pixel represents wheat in an agricultural field, then its neighboring pixels also represent wheat with high probability [24]. Any statistical model based on the training samples of these neighboring pixels must take into account these correlations, and BCS structure could be a reasonable assumption.

Like multivariate spatial data, analysis of multivariate time series data is important in many fields such as medicine, finance, engineering and environmental science. Modeling of multivariate time series data may be used in many decision making activities, especially the prediction of todays financial market, and BCS structure could be a reasonable assumption for analyzing these kinds of financial data. On the other hand, climate change is the biggest challenge that the world is facing in the century, and is a significant and emerging threat to public health. Proper modeling of the climate change data can be addressed in the framework of BCS structure, and predict the environmental impacts. For example, BCS structure is a valuable tool for explaining the behavior of glaciers, for forecasting future environmental issues under a different set of observed conditions, and how glaciers affect stream flow by taking into consideration the multivariate observations over time. Scientists are interested in collecting ice cores (cylinders of ice) and drill out ice cores from an ice sheet or glacier that have recrystallized and have trapped small air bubbles. These small bubbles of air contain a sample of the atmosphere—from these it is possible to measure directly the past concentration of gases including carbon dioxide and methane (m) in the atmosphere. These cores are continuous records providing scientists with year-by-year (u) information about past climate. Scientists analyze various components of cores, particularly trapped air bubbles, which reveal past atmospheric composition,