Journal of Statistical Planning and Inference 142 (2012) 2017–2031

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Estimation of parameters in a generalized GMANOVA model based on an outer product analogy and least squares

Jianhua Hu [a,*], Fuxiang Liu [a,b], Jinhong You [a]

[a] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, PR China
[b] Science College and Institute of Intelligent Vision and Image Information, China Three Gorges University, Yichang, Hubei 443002, PR China

ARTICLE INFO

Article history:
Received 2 December 2009
Received in revised form
6 September 2011
Accepted 6 January 2012
Available online 14 February 2012

Keywords:
Asymptotic normality
Generalized GMANOVA model
Least squares
Two-stage generalized least squares
Outer product least squares

ABSTRACT

This paper investigates estimation of parameters in a combination of the multivariate linear model and growth curve model, called a generalized GMANOVA model. Making analogy between the outer product of data vectors and covariance yields an approach to directly do least squares to covariance. An outer product least squares estimator of covariance (COPLS estimator) is obtained and its distribution is presented if a normal assumption is imposed on the error matrix. Based on the COPLS estimator, two-stage generalized least squares estimators of the regression coefficients are derived. In addition, asymptotic normalities of these estimators are investigated. Simulation studies have shown that the COPLS estimator and two-stage GLS estimators are alternative competitors with more efficiency in the sense of sample mean, standard deviations and mean of the variance estimates to the existing ML estimator in finite samples. An example of application is also illustrated.

1. Introduction

In a variety of areas, observations that occur in social studies, biological science, economics and medical research are measured over multiple time points on a particular characteristic to investigate temporal pattern of change on the characteristic. The observations of repeated measurements are usually analyzed by the growth curve model. The model, initiated by Potthoff and Roy (1964), was studied by many researchers, including Khatri (1966), Grizzle and Allen (1969), Rao (1965, 1987), Lee (1988), Lange and Laird (1989), Pan and Fang (2002), Kollo et al. (2007), Hu and Yan (2008) and among others.

Undoubtedly, the growth curve model is useful for analyzing repeated measurements. However, except for the index of groups and observing time points it is hard to take other compound covariates into account. In order to avoid this kind of awkward the extensions of the growth curve model are needed. One of the important extensions is a combination of the multivariate linear model and growth curve model, which has the following form:

$$\mathbf{Y} = X_1 \Theta_1 Z' + X_2 \Theta_2 + \mathcal{E}, \tag{1}$$

where $\mathbf{Y}$ is the observations matrix of the response consisting of p repeated measurements taken on n individuals, X_1 is an $n \times m$ treatment design matrix, Z is a $p \times q$ profile matrix, X_2 is an $n \times s$ compound covariate matrix, and Θ_1 and Θ_2 are

* Corresponding author.
E-mail addresses: frank.jianhuahu@gmail.com (J. Hu), liufuxiangst@gmail.com (F. Liu), johnyou@gmail.com (J. You).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2012.01.022

unknown $m\times q$ and $s\times p$ matrices of the regression coefficient parameters. We assume that observations on individuals are independent, so that the rows of the random error matrix $\mathcal{E}$ are independent and identically distributed by a general continuous type distribution $\mathcal{G}$ with mean zero and a common covariance matrix Σ of order p. When $\Theta_2=\mathbf{0}$ the model (1) reduces to the traditional growth curve model and when $\Theta_1=\mathbf{0}$ the model (1) becomes the multivariate linear model.

Several authors have studied the estimating problem for the model (1), including Chinchilli and Elswick (1985) on MLEs of parameters, Gupta and Kabe (2000) on likelihood ratio test, Bai (2005) on exact distributions of MLE of covariance, and Bai and Shi (2007) on exact distributions of MLEs of regression coefficient matrices. All these work needs an assumption of normal error distribution. Sometimes, errors being normal may be not true. When the error distribution is not normal, the estimations developed by above authors may result in the wrong statistical inference results. Therefore, it is essential to develop a free distribution statistical inference method for the model (1) without assumption of normality. Since two column spaces $\mathcal{C}(X_1)$ and $\mathcal{C}(X_2)$ are free of any inclusion relationship, the model (1) is not the special case of the extended growth curve model which was investigated in von Rosen (1989, 1991). The model (1) is not covered by the additive growth curve model with orthogonal design matrices, studied in Hu et al. (in press, 2011), either.

The organization of the paper is as follows. In Section 2, an analogy between outer product of a data vector and covariance is introduced. Based on the analogy, an auxiliary least squares model is explored to directly do least squares to covariance. The normal equations and the solution, called a COPLS estimator, are presented. In Section 3, the distribution of the COPLS estimator is obtained if normal distribution is imposed on the random error. The generalized least squares estimators of the regression coefficient matrices are obtained via taking the least squares estimator of covariance matrix as the first step estimate of Σ in Section 4. Both the consistency and asymptotic normality of the estimators derived in the previous sections are investigated under some certain conditions in Section 5. Simulation studies and an application are provided in Section 6. Finally, the brief concluding remarks are stated in Section 7.

2. An outer product least squares estimator of covariance based on analogy

For an $n\times p$ matrix $\mathbf{Y}$, write $\mathbf{Y}=[\mathbf{y}_1',\ldots,\mathbf{y}_n']'$, $\mathbf{y}_i'\in\mathbb{R}^p$, where $\mathbb{R}^p$ is the p-dimensional real space, and $\mathbf{y}=\mathrm{vec}(\mathbf{Y})$ denotes an np-dimensional vector $[\mathbf{y}_1,\ldots,\mathbf{y}_n]'$. Here the vec operator transforms a matrix into a vector by stacking the rows of the matrix one underneath another. $\mathcal{M}_{n\times p}$ denotes the set of all $n\times p$ matrices over real set $\mathbb{R}$ with trace inner product $\langle,\rangle$ and $\|\cdot\|$ denotes the trace norm on the set $\mathcal{M}_{n\times p}$. $\mathcal{S}_p$ denotes the set of all $p\times p$ symmetric matrices over real set $\mathbb{R}$. $\mathcal{N}_p$ denotes the set of all nonnegative definite matrices of order p. The Kronecker product $\otimes$ is defined as $A\otimes B=(a_{ij}B)$ with $A=(a_{ij})$ and $\mathrm{vec}(ABC)=(A\otimes C')\mathrm{vec}(B)$.

For the Gauss–Markov linear model $\mathbf{y}_{n\times 1}=X_{n\times m}\boldsymbol{\beta}+\boldsymbol{\epsilon}_{n\times 1}$ with $\mathrm{E}(\boldsymbol{\epsilon})=\mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon})=\sigma^2 I_n$, the ordinary least squares method is to find an m-dimensional vector $\widehat{\boldsymbol{\beta}}(\mathbf{y})$ in the real space $\mathbb{R}^m$ such that

$$\widehat{\boldsymbol{\beta}}(\mathbf{y})=\underset{\boldsymbol{\beta}\in\mathbb{R}^m}{\mathrm{argmin}}\|\mathbf{y}-X\boldsymbol{\beta}\|^2. \tag{2}$$

Equivalently, the ordinary least squares method takes the perpendicular projection $P_X\mathbf{y}$ of $\mathbf{y}$ as the least squares estimator of the expected value $\mathrm{E}(\mathbf{y})$, that is

$$\widehat{X\boldsymbol{\beta}(\mathbf{y})}=X\widehat{\boldsymbol{\beta}}(\mathbf{y})=P_X\mathbf{y}.$$

After finding the ols estimator $\widehat{\boldsymbol{\beta}}(\mathbf{y})$, the following statistic:

$$\widehat{\sigma}^2_{ols}(\mathbf{y})=\frac{1}{n-r}(\mathbf{y}-\widehat{X\boldsymbol{\beta}(\mathbf{y})})'(\mathbf{y}-\widehat{X\boldsymbol{\beta}(\mathbf{y})})=\frac{1}{n-r}\mathbf{y}'(I-P_X)\mathbf{y} \tag{3}$$

is constructed by summing the residual squares, where r is the rank of the design matrix X. The expression (3) is usually viewed as the ordinary least squares estimator of σ^2. This technique cannot extend to many complicated statistical models, e.g., the generalized GMANOVA model and others.

To overcome the drawback, a least squares problem directly to covariance will be deliberately designed in this section. We need the following concept of outer product.

The *outer product* over the np-dimensional real space $\mathbb{R}^{np}$ is defined as

$$\mathbf{a}\,\square\,\mathbf{b}=\mathbf{a}\mathbf{b}'=(a_ib_j)_{np\times np}$$

for any $\mathbf{a}=(a_1,\ldots,a_{np})',\mathbf{b}=(b_1,\ldots,b_{np})'\in\mathbb{R}^{np}$. The covariance of two np-dimensional random vectors $\mathbf{y}$ and $\mathbf{z}$ can be viewed as *a special outer product* $\square_c$ defined by

$$\mathbf{y}\,\square_c\,\mathbf{z}=\mathrm{Cov}(\mathbf{y},\mathbf{z})=(\mathrm{Cov}(y_i,z_j))_{np\times np}.$$

The outer product of the data vector in a random sample should contain the information of the behavior of the unknown parameters in covariance of the population. Therefore, we will consider using the outer products of a random sample to estimate covariance of its population.

Let $T=(X_1\otimes Z,X_2\otimes I)$ and $P_T=T(T'T)^-T'$, the orthogonal projection onto the column space $\mathcal{C}(T)$ of the matrix T, then $M=I-P_T$ is the orthogonal projection onto the orthogonal complement $\mathcal{C}(T)^\perp$ of $\mathcal{C}(T)$.