Journal of Statistical Planning and Inference 139 (2009) 3013 – 3026

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

A robust approach to joint modeling of mean and scale covariance for longitudinal data

Tsung-I. Lin[a,*], Yun-Jen Wang[b]

[a]Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan
[b]Graduate Institute of Finance, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 7 November 2007
Received in revised form
4 November 2008
Accepted 12 February 2009
Available online 3 March 2009

Keywords:
Covariance structure
Maximum likelihood estimates
Reparameterization
Robustness
Outliers
Prediction

ABSTRACT

In this paper, we propose a multivariate t regression model with its mean and scale covariance modeled jointly for the analysis of longitudinal data. A modified Cholesky decomposition is adopted to factorize the dependence structure in terms of unconstrained autoregressive and scale innovation parameters. We present three distinct representations of the log-likelihood function of the model and study the associated properties. A computationally efficient Fisher scoring algorithm is developed for carrying out maximum likelihood estimation. The technique for the prediction of future responses in this context is also investigated. The implementation of the proposed methodology is illustrated through two real-life examples and extensive simulation studies.

1. Introduction

In recent years, the method of joint modeling of mean and covariance on the general linear model with multivariate normal errors, called the normal joint modeling model (NJMM) hereafter, was heuristically introduced by Pourahmadi (1999, 2000). The key advantages of such normal-error models include the convenience in statistical interpretation and computational ease in parameter estimation. Yet it still exists several weaknesses. For instance, the assumption of normality for the error terms may be questionable in many practical situations when atypical points exist or the underlying data exhibit thick tails. A number of authors in the literature have used a more thick-tailed distribution, like the multivariate t distribution, in place of normal for robust estimation of general linear models (Zellner, 1976; Lange et al., 1989; He et al., 2004). Robust estimation for linear mixed models using the multivariate t distribution has been studied by Welsh and Richardson (1997) and Pinheiro et al. (2001), among others.

Specifically, a p-dimensional random vector $\boldsymbol{Y}$ is said to follow a multivariate t distribution with degrees of freedom (df) v, mean vector $\boldsymbol{\mu}$ and scale covariance matrix $\boldsymbol{\Sigma}$ if its probability density function is

$$f(\boldsymbol{Y}|\boldsymbol{\mu},\boldsymbol{\Sigma},v)=\frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(\pi v)^{p/2}}|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{(\boldsymbol{Y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})}{v}\right)^{-(v+p)/2}.$$

We shall use the notation $\boldsymbol{Y}\sim T_p(\boldsymbol{\mu},\boldsymbol{\Sigma},v)$ to denote that $\boldsymbol{Y}$ follows the above distribution. The multivariate t distribution has attracted considerable attention over the past 20–30 years. It has been applied in a wide variety of research fields, see Kotz and Nadarajah (2004) and the references therein.

* Corresponding author. Tel.: +886 4 22850420; fax: +886 4 22873028.
E-mail address: tilin@amath.nchu.edu.tw (T.-I. Lin).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2009.02.008

In this paper, we extend Pourahmadi's approach of joint mean–covariance parameterization to general linear models with the error term distributed according to a multivariate t distribution, also called the t joint modeling model (TJMM), as a robust approach to the analysis of longitudinal data.

Suppose that the repeated measurements of a continuous random variable are observed over time on each of m subjects. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^{\mathrm{T}}$ be the response vector for the ith subject measured at time points $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}}$, which are allowed unevenly spaced, and let the associate covariates $\boldsymbol{X}_i = [\boldsymbol{x}_{i1} \ldots \boldsymbol{x}_{ip}]$ be an $n_i \times p$ full rank design matrix.

The TJMM is defined as

$$\boldsymbol{Y}_i \sim T_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu) \quad (i = 1, \ldots, m), \tag{1}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^{\mathrm{T}} = \boldsymbol{X}_i \boldsymbol{\beta}$ is the mean response vector for subject i. Moreover, to ensure positive definiteness of $\boldsymbol{\Sigma}_i = [\sigma_{ij}]$, we reparameterize it via the modified Cholesky decomposition as

$$\boldsymbol{L}_i \boldsymbol{\Sigma}_i \boldsymbol{L}_i^{\mathrm{T}} = \boldsymbol{D}_i, \tag{2}$$

where $\boldsymbol{D}_i = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_{n_i}^2\}$ and $\boldsymbol{L}_i = [\ell_{jk}]$ is a unit lower triangular matrix with the (j, k)th entry being $-\phi_{jk}$. Obviously, $\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{L}_i^{\mathrm{T}} \boldsymbol{D}_i^{-1} \boldsymbol{L}_i$. The parameters ϕ_{jk} and σ_j^2 in $\boldsymbol{L}_i$ and $\boldsymbol{D}_i$ are referred to as the *autoregressive parameters* and *scale innovation variances* of $\boldsymbol{\Sigma}_i$, respectively. Note that such decomposition in (2) is unique and has several nice features. For a detailed discussion on the modified Cholesky decomposition, interested readers are referred to Pourahmadi (2001, Section 3.5).

Statistical interpretations for such reparameterization include (a) the below-diagonal entries of $\boldsymbol{L}_i$ are the negatives of the autoregressive parameters, namely $-\phi_{jk}$, in

$$\hat{Y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{jk}(Y_{ik} - \mu_{ik}),$$

which is the linear least-squares predictor of Y_{ij} based on its predecessors; (b) the diagonal entries of $\boldsymbol{D}_i$ are the scale innovation variances $\sigma_j^2 = c_\nu^{-1} \mathrm{var}(Y_{ij} - \hat{Y}_{ij})$, where $c_\nu = \nu/(\nu - 2)$.

To make the dimension of unconstrained parameters ϕ_{jk} and $\log \sigma_j^2$ more parsimonious, we model them using covariates in the spirit of Pourahmadi (1999), namely, for $j = 1, \ldots, n_i$ and $k = 1, \ldots, j-1$,

$$\phi_{jk} = \boldsymbol{z}_{jk}^{\mathrm{T}} \boldsymbol{\gamma}, \quad \log \sigma_j^2 = \boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{\lambda}, \tag{3}$$

where $\boldsymbol{z}_{jk}$ and $\boldsymbol{w}_j$ are $d \times 1$ and $q \times 1$ covariate vectors, which can usually be determined in terms of polynomial of measurement time t_{ij}'s with degrees of $d - 1$ and $q - 1$, respectively, and $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are $d \times 1$ and $q \times 1$ vectors of unknown parameters. Note that $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are assumed to be common for all $\boldsymbol{\Sigma}_i$'s for exhibiting the same covariance structure. In other words, $\boldsymbol{\Sigma}_i$ depends on i only through its dimension $n_i \times n_i$.

The rest of the paper is organized as follows. In the next section, we present three distinct representations of the log-likelihood function of TJMM and describe a Fisher scoring algorithm for the implementation of ML estimation. Section 3 is devoted to addressing the prediction issue. For illustration purposes, two real examples are presented in Section 4 and extensive simulation results are reported in Section 5. Finally, some concluding remarks are briefly summarized in Section 6, and the technical derivations are sketched in Appendices.

2. Computational aspects of parameter estimation

2.1. The log-likelihood function

For notational convenience, let $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} = \{r_{ij}\}_{j=1}^{n_i}$, $\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \boldsymbol{r}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{r}_i$ and $n = \sum_{i=1}^m n_i$ be the total number of observations. Denote by $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \nu)$ the population model parameters vector, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)$. Given independent observations $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)$, the log-likelihood function of $\boldsymbol{\alpha}$ corresponding to TJMM can be written in three distinct representations as follows:

$$\begin{aligned}
\ell(\boldsymbol{\alpha}|\boldsymbol{Y}) &= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\pi\nu) - \frac{1}{2}\sum_{i=1}^m \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^m (\nu + n_i)\log\left(1 + \frac{\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})}{\nu}\right) \\
&= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\pi\nu) - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^{n_i} \log \sigma_j^2 - \frac{1}{2}\sum_{i=1}^m (\nu + n_i)\log\left(1 + \frac{1}{\nu}\sum_{j=1}^{n_i} \eta_{ij}^2\right) \\
&= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\pi\nu) - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^{n_i} \boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{\lambda} - \frac{1}{2}\sum_{i=1}^m (\nu + n_i)\log\left(1 + \frac{(\boldsymbol{r}_i - \boldsymbol{Z}_i \boldsymbol{\gamma})^T \boldsymbol{D}_i^{-1} (\boldsymbol{r}_i - \boldsymbol{Z}_i \boldsymbol{\gamma})}{\nu}\right),
\end{aligned} \tag{4}$$