应用概率统计　第三十一卷
第六期　2015年12月

Chinese Journal of Applied Probability
and Statistics Vol.31 No.6 Dec. 2015

Joint Modeling of Mean-Covariance Structures Based on Partial Autocorrelation for Longitudinal Data *

Zhang Weiping*　Liu Yuting　Li Ruichao

(*Department of Statistics and Finance, University of Science and Technology of China, Hefei, 230026*)

Abstract

In this paper, we propose a joint mean-variance-correlation modeling approach for longitudinal studies. By applying partial autocorrelations, we obtain an unconstrained parametrization for the correlation matrix that automatically guarantees its positive definiteness, and develop a regression approach to model the correlation matrix of the longitudinal measurements by exploiting the parametrization. The proposed modeling framework is parsimonious, interpretable, and flexible for analyzing longitudinal data. Real data example and simulation support the effectiveness of the proposed approach.

Keywords: Correlation matrix, joint modeling, longitudinal data analysis, partial autocorrelation.

AMS Subject Classification: 62F12, 62F05.

§1. Introduction

In longitudinal studies, it is of fundamental importance to understand the dynamics in the mean function, variance function, and correlations. Simultaneously modeling the mean and covariance have recently attracted much attention, as for longitudinal study, the collected observations of the same subject are intrinsically correlated and misspecification of the correlation may result in a great loss of efficiency (Wang and Carey, 2003). Diggle et al. (2002) gave an excellent overview of various approaches to model the mean function for this type of data sets. However, modeling the covariance matrix is more challenging than modeling the mean as there are usually more parameters in the covariance matrix and the positive definiteness of the covariance matrix has to be assured. Pourahmadi

*The research was supported by the National Natural Science Foundation of China (11271347, 11171321), Natural Science Foundation of Anhui Province (1308085MA02) and the Fundamental Research Funds for the Central Universities.

*Corresponding author, E-mail: zwp@ustc.edu.cn.

Received July 26, 2014. Revised June 26, 2015.
doi: 10.3969/j.issn.1001-4268.2015.06.003

(1999, 2000) first introduced a modified Cholesky decomposition to factor the covariance matrix. An attractive property of this decomposition is to provide statistically meaningful unconstrained parametrization for the positive definite covariance matrix. The entries in this decomposition can be interpreted as autoregressive parameters and log innovation variances in a time series context. Regression models can then be applied to these entries in a manner similar to the mean models, thus permitting parsimonious characterization of the covariance structure just like the mean. See Pan and MacKenzie (2003), Ye and Pan (2006), Leng et al. (2010) and Zhang and Leng (2012) for recent developments and related discussions.

In this class of joint mean-covariance modeling approaches, however, the log innovation variances are not the conditional variances of the longitudinal response given the covariates. To extract the variance information, one must transform the respective decompositions back to the original covariance matrix that gives nontrivial interpretations with respect to the covariates. Similarly, additional steps are also needed to study the correlation matrix as an objective of interest in practice for quantifying the correlations among the longitudinal measurements. Specifically, a correlation matrix has unity diagonal entries, and must be positive definite with elements taking values between -1 and 1. A regression approach based on a direct Cholesky type decomposition of the correlation matrix can hardly satisfy the requirements, and hence it encounters great difficulty in this scenario. Therefore, extra effort and caution are required in practice to apply the aforementioned approaches for interpreting the features in the variance and covariations.

It is of great interest to develop an interpretable, efficient and flexible approach that targets directly at the variances and correlations in the longitudinal data. Daniels and Pourahmadi (2009) studied an unconstrained parametrization by exploiting the partial autocorrelation matrix. They parametrized a correlation matrix by the corresponding partial autocorrelation matrix, the resulting parameters can vary freely in the interval $(-1, 1)$. If needed, further transformation such as one involving Fisher's z transformation can give unconstrained parametrization in the entire real line $(-\infty, \infty)$. Such parametrization permits us to explore the variance information and interpret the variances with respect to the covariates directly, and quantify the correlations among the longitudinal measurements by partial autocorrelation matrix. However, they only focused on the Bayesian approach for the inference of parameters rather than studying the parsimonious maximum likelihood approach.

In this paper, by applying partial autocorrelations, we obtain an unconstrained parametrization for the correlation matrix that automatically guarantees its positive definite-

ness, and develop a parsimonious regression approach to model the correlation matrix of the longitudinal measurements. The proposed modeling framework is parsimonious, interpretable and flexible for analyzing longitudinal data. The maximum likelihood approach for parameters estimation is then studied. The rest of this paper is organized as follows. Section 2 elaborates the proposed joint modeling approach, the computational algorithm and its theoretical properties. We provide extensive numerical examples by applying our method to real data analysis, and conduct simulation in Section 3. The numerical results confirm the attractiveness of the new joint modeling approach. We conclude this paper by summarizing the main findings and outlining future research in Section 4. All technical proofs are relegated to the Appendix.

§2. The Model and the Estimating Method

Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})'$ be the m repeated measurements at time points $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{im})'$ on the ith subject $(i = 1, 2, \ldots, n)$. In more general setting, t_{ij} does not have to be time, but can be any time-dependent covariate being modeled parametrically. Let $\boldsymbol{x}_{ij}$ be p-vector covariates and $\boldsymbol{x}_i' = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im})$, in this paper we assume $\boldsymbol{y}_i|\boldsymbol{t}_i, \boldsymbol{x}_i \sim N_m(\mu_i, \Sigma_i)$ with $\mu_i = \mathsf{E}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{t}_i)$ and $\mathsf{Var}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{t}_i) = \Sigma_i = D_i R_i D_i$ with $D_i = \mathrm{diag}(\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{im})$ a diagonal matrix containing the marginal standard deviations of y_{ij} and $R_i = (\rho_{ijk})$ the correlation matrix.

The partial autocorrelation between y_{ij} and y_{ik} $(j < k)$, π_{ijk}, is the correlation between the two variables given the intervening variables $(y_{i(j+1)}, y_{i(j+2)}, \ldots, y_{i(k-1)})$. Let Π_i be the upper-triangular matrix with elements π_{ijk} $(j, k = 1, 2, \ldots, m)$. It can be show the relationship between $R_i = (\rho_{ijk})$ and $\Pi_i = (\pi_{ijk})$ (Anderson, 1984; Section 2.5) as the follows

$$\pi_{ij(j+1)} = \rho_{ij(j+1)}, \tag{2.1}$$

$$\pi_{ijk} = r_{i1}^{-1/2} r_{i2}^{-1/2} [\rho_{ijk} - \boldsymbol{r}_{i1}'(j,k) \boldsymbol{R}_{i3}(j,k)^{-1} \boldsymbol{r}_{i2}(j,k)], \qquad j - k > 1, \tag{2.2}$$

where $\boldsymbol{r}_{i1}'(j,k) = (\rho_{ij(j+1)}, \rho_{ij(j+2)}, \ldots, \rho_{ij(k-1)})$, $\boldsymbol{r}_{i2}'(j,k) = (\rho_{ik(j+1)}, \rho_{ik(j+2)}, \ldots, \rho_{ik(k-1)})$, and $\boldsymbol{R}_{i3}(j,k)$ is the sub-correlation matrix of R_i corresponding to the variables $(y_{i(j+1)}, y_{i(j+2)}, \ldots, y_{i(k-1)})$. The scalar r_{il} $(l = 1, 2)$ are $r_{il} = 1 - \boldsymbol{r}_{il}'(j,k) \boldsymbol{R}_{i3}(j,k)^{-1} \boldsymbol{r}_{il}(j,k)$. The partial autocorrelation coefficient π_{ijk} can be defined equivalently as the correlation between y_{ij} and y_{ik} after correcting for $\widetilde{\boldsymbol{y}}_{ij:k}' = (y_{i(j+1)}, y_{i(j+2)}, \ldots, y_{i(k-1)})$, that is $\pi_{ijk} = \mathrm{corr}(y_{ij} - \boldsymbol{b}_j' \widetilde{\boldsymbol{y}}_{ij:k}, y_{ik} - \boldsymbol{b}_k' \widetilde{\boldsymbol{y}}_{ij:k})$ with $\boldsymbol{b}_j' \widetilde{\boldsymbol{y}}_{ij:k}$ and $\boldsymbol{b}_k' \widetilde{\boldsymbol{y}}_{ij:k}$ the linear least squares predictors of y_{ij} and y_{ik} given $\widetilde{\boldsymbol{y}}_{ij:k}$.

It is clear from (2.1)-(2.2) that the mapping from R_i to Π_i is invertible. By inverting the previous operations recursively over increasing lag $k-j$, one obtains the correlation matrix by

$$\rho_{ij(j+1)} = \pi_{ij(j+1)}, \tag{2.3}$$

$$\rho_{ijk} = \boldsymbol{r}'_{i1}(j,k)\boldsymbol{R}_{i3}(j,k)^{-1}\boldsymbol{r}_{i2}(j,k) + r_{i1}^{1/2}r_{i2}^{1/2}\pi_{ijk}, \qquad \text{for } k-j>1. \tag{2.4}$$

The key advantage in using partial autocorrelation is that parameters are unconstrained and can automatically guarantee the positive definiteness of correlation matrix (Joe, 2006). It is clear that for the correlation matrix R_i, the subset of values in $(-1,1)$ that ρ_{ijk} can take satisfying the positive definite constraint is determined by the configuration of the other elements of R_i. Rousseeuw and Molenberghs (1994) gave a geometric interpretation of this phenomenon. While for the partial autocorrelation, each π_{ijk} can vary freely in the interval $(-1,1)$ and takes values in the entire real line after using Fisher's z transformation, regardless of the choice of the remaining $\pi's$. Therefore this mapping suggests a practical alternative by avoiding the complication of the positive definite constraint, while providing easily interpretable parameters (Joe, 2006). Thus, we are free to characterize these parameters via regression as functions of some covariates. In practice, such a rationale can be initially assessed by examining empirical variances and partial autocorrelations from the observed longitudinal data. For a balanced longitudinal study such as the cattle example in Section 3.1, an initial version of the partial autocorrelation π_{ijk} can be obtained from the empirical correlation matrix of the standardized residuals after a mean-variance model fitting. By examining the plot of those π_{ijk} after using Fisher's z transformation against the time lag and logarithm of σ_{ij} against time in Figure 1 of following example, we clearly observe a curvature that supports some functional associations. From there, appropriate models can be used to describe such a curvature.

Example 1 We assume a multivariate normal distribution with common mean and an 11×11 covariance matrix for the treatment group A with $n=30$ animals as Pourahmadi (1999). Figure 1 shows the sample partial autocorrelation coefficients after using Fisher's z transformation and logarithm of sample standard deviations. Clearly, these scatter plots indicate some functional relationships approximately. The partial autocorrelation and logarithm of standard deviation can be well explained by polynomials in time lag and time, respectively. Section 3.1 gives more detailed discussion.

Motivated by above considerations and cattle data example, we propose a joint regression model for the mean, the partial autocorrelations and the marginal standard deviations,

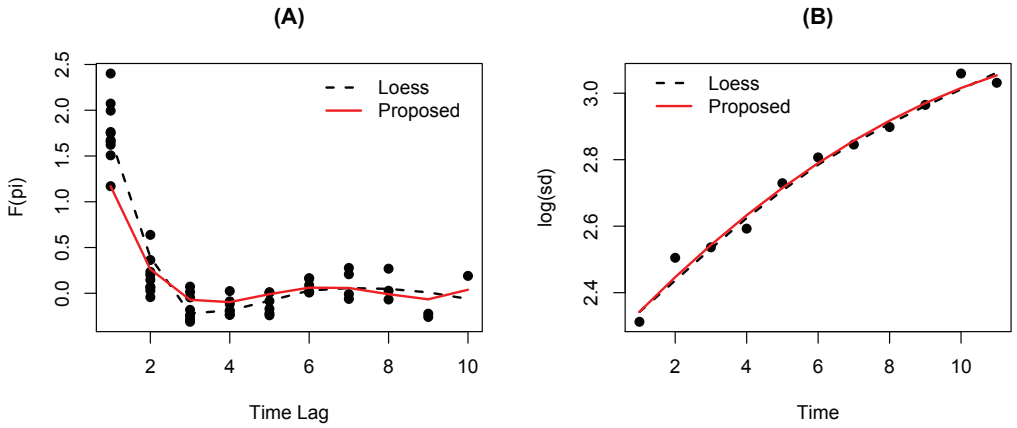

Figure 1 Sample regressograms for the cattle data. (A) The sample partial autocorrelation coefficients after using Fisher's z transformation and the fitted curve again time lag; (B) logarithm of sample standard deviations and the fitted curve against time.

as

$$g(\mu_{ij}) = \boldsymbol{x}'_{ij}\beta, \qquad F(\pi_{ijk}) = \boldsymbol{w}'_{ijk}\gamma, \qquad \log(\sigma_{ij}) = \boldsymbol{z}'_{ij}\lambda. \tag{2.5}$$

Such modeling approach is also used by many other authors, for example, see Pourahmadi (1999, 2000) and Pan and MacKenzie (2003). Here $g(\cdot)$ and $F(\cdot)$ are monotone and differentiable known link functions, which can give unconstrained parametrization in the entire real line $(-\infty, \infty)$. We take F as the Fisher's z transformation in this paper. This facilitates a convenient and flexible modeling device for the correlation matrix. $\boldsymbol{x}_{ij}, \boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$ are the $p \times 1$, $q \times 1$ and $d \times 1$ vectors of covariates, respectively. The covariates $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are those used in regression analysis, while $\boldsymbol{w}_{ijk}$ is usually taken as a polynomial of time difference $t_{ij} - t_{ik}$ or that of time dependent covariates. Thus, the difference between observations at two continuous times t_{ij} and t_{ik} is reflected in $\boldsymbol{w}_{ijk}$ that explains the covariation between y_{ij} and y_{ik}.

Let $\theta = (\beta', \gamma', \lambda')'$, write the minus twice the log-likelihood function $l(\theta)$, up to a constant, as

$$\begin{aligned} -2l(\theta) &= \sum_{i=1}^{n} \log|\Sigma_i| + \sum_{i=1}^{n} (\boldsymbol{y}_i - \mu_i)^{\tau} \Sigma_i^{-1} (\boldsymbol{y}_i - \mu_i) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \log(\sigma_{ij}^2) + \sum_{i=1}^{n} \log|R_i(\Pi)| + \sum_{i=1}^{n} (\boldsymbol{y}_i - \mu_i)^{\tau} D_i^{-1} R_i(\Pi)^{-1} D_i^{-1} (\boldsymbol{y}_i - \mu_i), \end{aligned} \tag{2.6}$$

where $R_i(\Pi)$ denotes correlation matrix corresponding to the partial autocorrelation Π_i.

Remark 1 Daniels and Pourahmadi (2009) shows that

$$|\Sigma_i| = \prod_{l=1}^{m} \sigma_{il}^2 \prod_{j=2}^{m} \prod_{k=1}^{j-1} (1 - \pi_{ijk}^2),$$

thus the $\log |R_i(\Pi)|$ term can be expressed to $\sum_{j=2}^{m} \sum_{k=1}^{j-1} \log(1 - \pi_{ijk}^2)$.

By taking partial derivatives of $l(\theta)$ with respect to these parameters respectively, the maximum likelihood estimating equations become

$$\begin{aligned} U_1(\beta;\gamma,\lambda) &= \sum_{i=1}^{n} \boldsymbol{x}_i' \Delta_i \Sigma_i^{-1} (\boldsymbol{y}_i - \mu(\boldsymbol{x}_i\beta)) = 0, \\ U_2(\gamma;\beta,\lambda) &= \frac{1}{2} \sum_{i=1}^{n} \sum_{j,k} [R_i^{-1} D_i^{-1} (S_i - \Sigma_i) D_i^{-1} R_i^{-1}]_{jk} \frac{\partial \rho_{ijk}}{\partial \gamma} = 0, \\ U_3(\lambda;\beta,\gamma) &= \sum_{i=1}^{n} \boldsymbol{z}_i' (\mathrm{diag}\{R_i^{-1} D_i^{-1} S_i D_i^{-1}\} - I_m) \mathbf{1}_m = 0, \end{aligned} \tag{2.7}$$

where $\Delta_i = \Delta_i(\boldsymbol{x}_i\beta) = \mathrm{diag}(\dot{g}^{-1}(x_{ij}'\beta), \ldots, \dot{g}^{-1}(x_{im}'\beta))$, $\dot{g}^{-1}(\cdot)$ is the derivative of the inverse of the link function $g^{-1}(\cdot)$ and we have used the notation $\mu(\cdot) = g^{-1}(\cdot)$; $S_i = (\boldsymbol{y}_i - \mu_i)(\boldsymbol{y}_i - \mu_i)'$; $[M]_{jk}$ is the (j,k)th element of matrix M; $\mathbf{1}_m$ is a vector of 1's and I_m is the identity matrix of size m. The partial derivatives $\{\partial \rho_{ijk}/\partial \gamma\}$ are given in the Appendix.

The solutions of β, γ and λ satisfy the equations in (2.7). These parameters can be solved iteratively by fixing the other parameters. An application of the quasi-Newton type algorithm on equation (2.7) directly yields the numerical solutions for these parameters. However, computing the explicit form of the expectation of the Hessian or second-order derivative matrix with respect to $(\gamma', \lambda')'$ analytically is cumbersome, see some computation in the Appendix. Thus, for all our computations here, we use the analytic form of the first-order derivatives $(U_1', U_2', U_3')'$ in the following algorithm, but compute the second-order derivatives $\nabla^2_{\gamma,\lambda} l(\beta,\gamma,\lambda)$ numerically. More specifically, the algorithm works as follows.

1. Initialize the parameters as $\beta^{(0)}$, $\gamma^{(0)}$ and $\lambda^{(0)}$. Set $k = 0$.
2. Compute Σ_i using $\gamma^{(k)}$ and $\lambda^{(k)}$. Update β as

$$\beta^{(k+1)} = \beta^{(k)} + [I_{11}^{-1}(\theta) U_1(\beta;\gamma,\lambda)]|_{\beta=\beta^{(k)}}, \tag{2.8}$$

where $I_{11} = -\mathsf{E}[\partial^2 l(\theta)/\partial\beta\partial\beta'] = \sum_{i=1}^{n} \boldsymbol{x}_i' \Delta_i \Sigma_i^{-1} \Delta_i \boldsymbol{x}_i$.

3. Given $\beta = \beta^{(k+1)}$, update γ and λ using

$$\begin{pmatrix} \gamma^{(k+1)} \\ \lambda^{(k+1)} \end{pmatrix} = \begin{pmatrix} \gamma^{(k)} \\ \lambda^{(k)} \end{pmatrix} - \left[(\mathsf{E}\nabla^2_{\gamma,\lambda} l(\beta,\gamma,\lambda))^{-1} \begin{pmatrix} U_2(\gamma;\beta,\lambda) \\ U_3(\lambda;\beta,\gamma) \end{pmatrix} \right] \Bigg|_{\gamma=\gamma^{(k)},\lambda=\lambda^{(k)}}. \tag{2.9}$$

4. Set $k \leftarrow k+1$ and repeat Steps 2-3 until a pre-specified convergence criterion is met.

A natural starting value for β is to use identity matrices for Σ_i's in (2.7). γ and λ can be initiated by their estimates based on empirical covariance. It is not difficult to see that these initial estimates are $\sqrt{n}$-consistent. From the theoretical analysis in Theorem 1 in Section 2.1 and the proofs in the Appendix, the log-likelihood function is asymptotically convex around a small neighborhood of the true parameters. To ensure that the optimum is global, we may try multiple initial values for the parameters. For our data analysis and simulation studies, the algorithm is quite stable and convergence was usually obtained within several iterations.

2.1 Asymptotic Properties

Since we use maximum likelihood for estimation, the resulting estimators are efficient. To formally establish the theoretical properties of the estimates, we impose the following regularity conditions.

Condition A1: The dimensions p, q and d of covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$ are fixed; $n \to \infty$ and m is bounded.

Condition A2: The parametric space Θ is a compact subset of $\mathbb{R}^{p+q+d}$, and the parameter value $\theta_0 = (\beta_0', \gamma_0', \lambda_0')'$ is in the interior of the parameter space Θ.

Condition A1 and A2 are standard in longitudinal data analysis. The asymptotic property of the maximum likelihood estimation involves the negative of the expected Hessian matrix, $I(\theta) = -\mathsf{E}(\partial^2 l/\partial\theta\partial\theta')$ with $\theta = (\beta', \gamma', \lambda')'$. Here the expectation is conditional on the covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$.

Condition A3: When $n \to \infty$, $I(\theta_0)/n$ converges to a positive definite matrix $\mathscr{I}(\theta_0)$.

Condition A3 is standard in regression analysis. Formally, we have the following asymptotic results for the maximum likelihood estimates of the mean coefficients β, the moving average coefficients γ and the innovation coefficients λ.

Theorem 2.1 If $n \to \infty$ and regularity conditions A1-A3 hold, we have that: (a) the maximum likelihood estimator $(\widehat{\beta}', \widehat{\gamma}', \widehat{\lambda}')'$ is strongly consistent for the true value $(\beta_0', \gamma_0', \lambda_0')'$; and (b) $(\widehat{\beta}', \widehat{\gamma}', \widehat{\lambda}')'$ is asymptotically normally distributed as

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to N[0, \{\mathscr{I}(\theta_0)\}^{-1}].$$

It is shown in the Appendix that $\widehat{\beta}$ and $(\widehat{\gamma}', \widehat{\lambda}')'$ are asymptotically independent. The asymptotic covariance matrix $\mathscr{I}(\theta_0)$ can be estimated by using Bootstrap method or numerical second-order derivatives of likelihood evaluated at the $\widehat{\theta}$.

§3. Data Analysis and Simulations

3.1 Cattle Data

We first apply our approach to a balanced longitudinal data set in Kenward (1987), where cattle were assigned randomly to two treatment groups A and B, and their weights were measured 11 times over a 133-day period. As in Pourahmadi (2000) and Pan and MacKenzie (2003), we focus on the 30 animals in group A using a saturated mean model with 11 parameters. By examining the sample partial autocorrelation coefficients after using Fisher's z transformation versus the time lag between measurements in Figure 1, we see a clear curvature pattern that can be reasonably captured by a polynomial. Figure 1 also indicates a curvature pattern by examining the log sample variances versus the time of measurements. Thus, we propose two polynomials of time that defines $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$ for modeling the partial autocorrelations and the log-variances.

$$F(\pi_{ijk}) = \gamma_0 + \gamma_1(t_j - t_k) + \cdots + \gamma_4(t_j - t_k)^4;$$
$$\log(\sigma_{ij}) = \lambda_0 + \lambda_1 t_j + \lambda_2 t_j^2.$$

Using the algorithm (2.8) and (2.9), we obtain the maximum likelihood estimates, $\widehat{\gamma}_0 = 2.982_{0.346}$, $\widehat{\gamma}_1 = -2.395_{0.394}$, $\widehat{\gamma}_2 = 0.651_{0.152}$, $\widehat{\gamma}_3 = -0.073_{0.023}$, $\widehat{\gamma}_4 = 0.003_{0.001}$ and $\widehat{\lambda}_0 = 2.231_{1.761}$, $\widehat{\lambda}_1 = 0.115_{0.563}$, $\widehat{\lambda}_2 = -0.004_{0.046}$, the standard deviation in the parentheses are estimated using Bootstrap method that resamples with respect to the subjects in Kenward's cattle data 100 times. Figure 1 shows the fitted lines by Loess and our models which basically coincide with each other.

3.2 Simulation Studies

In this section we investigate the finite sample performance of the proposed estimation and inference methods with Monte Carlo simulation studies. We generate 500 data sets respectively, each consisting of n subjects. We consider the sample size $n = 50$, 150 and 250 respectively.

Study 1: This is to demonstrate the asymptotic properties in Section 2.1. The data sets are generated from the model

$$\begin{aligned} &y_{ij} = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + e_{ij}, \qquad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m), \\ &F(\pi_{ijk}) = \gamma_0 + w_{ijk1}\gamma_1 + w_{ijk2}\gamma_2, \\ &\log(\sigma_{ij}) = \lambda_0 + z_{ij1}\lambda_1 + z_{ij2}\lambda_2, \end{aligned} \tag{3.1}$$

where $m = 6$, and the measurement time $t_{ij} = 1, 2, \ldots, 6$. The true parameters $\beta = (1, -0.5, 0.5)'$, $\gamma = (0.5, -1, 0.5)'$ and $\lambda = (0.5, -0.5, 0)'$. The covariate $\boldsymbol{x}_{ij} = (1, x_{ij1}, x_{ij2})'$ is generated from a multivariate normal distribution with mean zero, marginal variance 1 and AR(1) correlation with the parameter equals to 0.5. We take $z_{ij} = (1, t_{ij}, t_{ij}^2)'$, and $w_{ijk} = (1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2)'$.

Table 1 shows the accuracy of the estimated parameters in terms of their mean (Mean), mean absolute biases (MAB) and standard deviations. It is clearly that our estimating methods literally yield unbiased estimates for the parameters, and all the biases are relatively small especially when n is large. Additionally, to evaluate the inference procedure, we compare the sample standard deviation (SD) of 500 parameter estimates to the sample average of 500 standard errors (SE) using numerical Hessian. They are close, especially for larger n. And we should note that the mapping form correlation to partial autocorrelation is nonlinear and complex, the standard deviations for $\widehat{\gamma}$ are reasonably larger than that for $\widehat{\beta}$ and $\widehat{\lambda}$ according to the delta theorem.

Table 1 Simulation results over 500 replications

	$n = 50$			$n = 150$			$n = 250$		
	Mean	MAB	SD_{SE}	Mean	MAB	SD_{SE}	Mean	MAB	SD_{SE}
β_0	1.003	0.063	$0.079_{0.057}$	1.001	0.037	$0.045_{0.033}$	0.998	0.029	$0.036_{0.026}$
β_1	-0.495	0.100	$0.124_{0.080}$	-0.504	0.051	$0.064_{0.047}$	-0.504	0.045	$0.056_{0.036}$
β_2	0.495	0.099	$0.122_{0.081}$	0.505	0.051	$0.063_{0.048}$	0.505	0.042	$0.052_{0.037}$
γ_0	0.474	0.541	$0.668_{0.486}$	0.508	0.317	$0.407_{0.288}$	0.490	0.242	$0.303_{0.216}$
γ_1	-1.000	2.023	$2.516_{1.830}$	-1.056	1.188	$1.499_{1.077}$	-0.995	0.907	$1.117_{0.800}$
γ_2	0.545	1.786	$2.229_{1.635}$	0.546	1.044	$1.306_{0.951}$	0.509	0.791	$0.970_{0.706}$
λ_0	0.476	0.144	$0.185_{0.137}$	0.493	0.083	$0.104_{0.083}$	0.493	0.063	$0.079_{0.060}$
λ_1	-0.448	0.564	$0.717_{0.532}$	-0.481	0.322	$0.404_{0.331}$	-0.486	0.244	$0.307_{0.236}$
λ_2	-0.043	0.470	$0.594_{0.447}$	-0.022	0.271	$0.338_{0.278}$	-0.011	0.203	$0.258_{0.197}$

Study 2: By decomposing the covariance matrix, Pourahmadi (2000) proposed an unconstrained joint mean-covariance modeling (MCD) approach based on normal likelihood approach. To compare the proposed approach with this kind of mean-covariance modeling approach, we define the following error measurements

$$\|\widehat{\mu}_d\| = \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i^{\mathrm{T}}(\widehat{\beta} - \beta_0)\|, \qquad \text{KL} = \frac{1}{n}\sum_{i=1}^{n}\text{KL}_i(f_{i1}|f_{i0}),$$

$$\|\widehat{\Sigma}_d\| = \frac{1}{n}\sum_{i=1}^{n}\|\widehat{\Sigma}_i - \Sigma_{0i}\|, \qquad \|\widehat{R}_d\| = \frac{1}{n}\sum_{i=1}^{n}\|\widehat{R}_i - R_{0i}\|,$$

where KL is the Kullback-Leibler divergence between a fitted model $f_{i1} = N(\widehat{\mu}_i, \widehat{\Sigma}_i)$ and the true model $f_{0i} = N(\mu_{0i}, \Sigma_{0i})$ for the ith subject. We consider the following three cases under sample sizes $n = 50$, 100. (I) We take the model (3.1) in Study 1 to generate data sets. In this case, the covariance model for the MCD approach is misspecified. (II) A similar model structure as in Case I is implemented by changing $F(\pi_{ijk})$ and $\log(\sigma_{ij})$ in (3.1) to the generalized autoregressive coefficient and log-innovation, respectively. In this case, the variance function and correlation structure in our approach are misspecified. (III) To compare these two methods when models are misspecified for both approaches, we take the same mean model as in Case I with the marginal variance $\sigma^2(t) = 0.5\mathrm{e}^t$ and ARMA(1,1) correlation structure

$$\mathrm{corr}(\epsilon_t, \epsilon_s) = \gamma\rho^{|t-s|},$$

for $t \neq s$. We consider $\gamma = 0.85$ and $\rho = 0.6$ corresponding to moderately correlated errors. In this case, both approaches use misspecified models for the covariance, since this correlation structure does not exactly correspond to either decomposition. The best these two approaches can do is to capture some signals in this correlation with their respective model specifications.

Table 2 Comparison between the proposed method and MCD method over 500 replications

	Results for PCA				Results for MCD			
n	$\|\widehat{\mu}_d\|$	$\|\widehat{\Sigma}_d\|$	$\|\widehat{R}_d\|$	KL	$\|\widehat{\mu}_d\|$	$\|\widehat{\Sigma}_d\|$	$\|\widehat{R}_d\|$	KL
Case I								
50	$0.310_{0.153}$	$1.055_{0.422}$	$0.387_{0.177}$	$0.112_{0.067}$	$0.319_{0.155}$	$1.474_{0.297}$	$0.572_{0.150}$	$0.175_{0.061}$
100	$0.223_{0.106}$	$0.730_{0.302}$	$0.256_{0.125}$	$0.050_{0.028}$	$0.228_{0.110}$	$1.255_{0.201}$	$0.486_{0.102}$	$0.116_{0.028}$
Case II								
50	$0.261_{0.126}$	$1.757_{0.255}$	$0.775_{0.181}$	$0.385_{0.081}$	$0.246_{0.115}$	$0.965_{0.409}$	$0.386_{0.168}$	$0.121_{0.060}$
100	$0.178_{0.096}$	$1.638_{0.197}$	$0.720_{0.131}$	$0.307_{0.043}$	$0.164_{0.088}$	$0.658_{0.281}$	$0.259_{0.120}$	$0.055_{0.027}$
Case III								
50	$0.239_{0.141}$	$1.896_{0.307}$	$1.459_{0.204}$	$0.645_{0.091}$	$0.252_{0.148}$	$1.984_{0.234}$	$1.474_{0.183}$	$0.836_{0.078}$
100	$0.169_{0.100}$	$1.865_{0.229}$	$1.441_{0.136}$	$0.570_{0.040}$	$0.180_{0.106}$	$1.943_{0.176}$	$1.455_{0.126}$	$0.768_{0.042}$

Table 2 shows the average of estimated error measurements over 500 replications and their sample standard deviations (in subscripts). It is not surprising from Table 2 that our approach substantially outperforms the alternative MCD approach in all the error

measurements under Case I, since the data are generated from our model. The MCD approach is recommended when the true model follows the modified Cholesky decomposition as seen by Case II. However, when the true model does not correspond to either of the competing models, our method still yields acceptable results in terms of the norms of the bias and the KL divergence, and performs much better than the Cholesky decomposition.

§4. Conclusion

We have proposed a joint maximum likelihood approach for modeling the mean, the variance and the correlation in longitudinal data analysis. Our approach permits unconstrained parametrization, fast computation and easy interpretation of the parameters. Unlike previous approaches, this approach targets directly at correlations and variances, and provides general form of the covariance structure. Our approach can also handle datasets observed at irregular time and highly unbalanced. In addition, the current regression model is fully parametric and based on likelihood. When nonlinearity arises, more flexible models may be called for.

Acknowledgements The authors are grateful to the editor and the anonymous referees for their helpful comments and suggestions, which have led to a substantially improved paper.

Appendix

A.1 The Derivative of the Correlation Coefficient

It is trivial for the case that $j = k$ and $k = j + 1$, for case of $j \neq k$, we can obtain $\partial\rho_{ijk}/\partial\gamma$ by using (2.3) and (2.4), the sth element of derivative $\partial\rho_{ijk}/\partial\gamma$,

$$\begin{aligned}
&\frac{\partial\rho_{ijk}}{\partial\gamma_s}\\
&= \frac{\partial\boldsymbol{r}'_{i1}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i2} + \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{R}_{i3}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i2} + \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{r}_{i2}}{\partial\gamma_s}\\
&\quad - \frac{1}{2}\pi_{ijk}\Big\{\frac{\partial\boldsymbol{r}'_{i1}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i1} + \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{R}_{i3}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i1} + \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{r}_{i1}}{\partial\gamma_s}\Big\}r_{i1}^{-1/2}r_{i2}^{1/2}\\
&\quad - \frac{1}{2}\pi_{ijk}\Big\{\frac{\partial\boldsymbol{r}'_{i3}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i3} + \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{R}_{i3}}{\partial\gamma_s}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i3} + \boldsymbol{r}'_{i3}(\boldsymbol{R}_{i3})^{-1}\frac{\partial\boldsymbol{r}_{i3}}{\partial\gamma_s}\Big\}r_{i1}^{1/2}r_{i2}^{-1/2}\\
&\quad + [1 - \boldsymbol{r}'_{i1}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i1}]^{1/2}[1 - \boldsymbol{r}'_{i2}(\boldsymbol{R}_{i3})^{-1}\boldsymbol{r}_{i2}]^{1/2}w_{ijk,s}. \qquad (4.1)
\end{aligned}$$

A.2 The Expectation of the Hessian

The computation of I_{11} is trivial. Since Σ only depends on γ and λ, it is easy to see that

$$I_{12}(\theta) = -\mathsf{E}\Big(\frac{\partial^2 l}{\partial\beta\partial\gamma'}\Big) = -\mathsf{E}\Big[\sum_{i=1}^{n} \boldsymbol{x}_i'\Delta_i\frac{\partial\Sigma_i^{-1}}{\partial\gamma'}\{\boldsymbol{y}_i - \mu(\boldsymbol{x}_i\beta)\}\Big] = 0.$$

Similarly $I_{13}(\theta) = 0$. For I_{33}, it is easy to see that

$$\begin{aligned} I_{33} &= -\mathsf{E}\Big(\frac{\partial^2 l}{\partial\lambda\partial\lambda'}\Big) = -\mathsf{E}\frac{\partial U_2}{\partial\lambda'} = \mathsf{E}\sum_{i=1}^{n}\frac{\partial}{\partial\lambda'}[\boldsymbol{z}_i'(\mathrm{diag}\{R_i^{-1}D_i^{-1}S_iD_i^{-1}\}\mathbf{1}_m] \\ &= 2\sum_{i=1}^{n}\boldsymbol{z}_i'(I_m + \Sigma_i^{-1}\circ\Sigma_i)\boldsymbol{z}_i, \end{aligned} \tag{4.2}$$

where $\circ$ represents the Hadamard product.

But computing I_{22} and I_{23} analytically are cumbersome,

$$\begin{aligned} I_{22}(\theta) &= -\mathsf{E}\Big(\frac{\partial^2 l}{\partial\gamma\partial\gamma'}\Big) = -\mathsf{E}\frac{\partial U_2}{\partial\gamma'} \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k}\frac{\partial\rho_{ijk}}{\partial\gamma}\mathsf{E}\Big\{\frac{\partial}{\partial\gamma'}[R_i^{-1}D_i^{-1}(S_i-\Sigma_i)D_i^{-1}R_i^{-1}]_{jk}\Big\}, \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} I_{23}(\theta) &= -\mathsf{E}\Big(\frac{\partial^2 l}{\partial\gamma\partial\lambda'}\Big) = -\mathsf{E}\frac{\partial U_2}{\partial\lambda'} \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k}\frac{\partial\rho_{ijk}}{\partial\gamma}\mathsf{E}\Big\{\frac{\partial}{\partial\lambda'}[R_i^{-1}D_i^{-1}(S_i-\Sigma_i)D_i^{-1}R_i^{-1}]_{jk}\Big\}. \end{aligned} \tag{4.4}$$

A.3 The Proof of Theorem 1

Proof The proof is essentially the same as that of Theorem 1 in Pourahmadi (2000) [17] and Theorem 1 and 2 in Chiu et al. (1996) [4].

(a) Let $l_i = \log f_i(\boldsymbol{y}_i,\theta)$, where f_i is the probability density function of $N_m(\mu_i,\Sigma_i)$ $(i = 1,2,\ldots,n)$. Then ignoring the constant $m\log(2\pi)/2$, we obtain that

$$l_i = -\frac{1}{2}\log(|\Sigma_i|) - \frac{1}{2}\{\boldsymbol{y}_i - \mu(\boldsymbol{x}_i\beta)\}'\Sigma_i^{-1}\{\boldsymbol{y}_i - \mu(\boldsymbol{x}_i\beta)\}.$$

Thus the mean and the variance of l_i when $\theta = \theta_0$ are respectively

$$\begin{aligned} \mathsf{E}_0(l_i) &= -\frac{1}{2}\log(|\Sigma_i|) - \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\Sigma_{0i}) - \frac{1}{2}\{\mu(\boldsymbol{x}_i\beta) - \mu(\boldsymbol{x}_i\beta_0)\}'\Sigma_i^{-1}\{\mu(\boldsymbol{x}_i\beta) - \mu(\boldsymbol{x}_i\beta_0)\}, \\ \mathsf{Var}_0(l_i) &= \frac{1}{2}[\mathrm{tr}(\Sigma_i^{-1}\Sigma_{0i})^2 + 2\{\mu(\boldsymbol{x}_i\beta) - \mu(\boldsymbol{x}_i\beta_0)\}'\Sigma_i^{-1}\Sigma_{0i}\Sigma_i^{-1}\{\mu(\boldsymbol{x}_i\beta) - \mu(\boldsymbol{x}_i\beta_0)\}], \end{aligned}$$

where $\Sigma_i = D_i R_i D_i^{\mathsf{T}}$ and $\Sigma_{0i} = D_{0i} R_{0i} D_{0i}^{\mathsf{T}}$. It follows from the compactness of the parameter space and boundedness of the covariates that $\mathsf{Var}_0(l_i) \leq K$, for all i where K is a constant. Therefore by Kolmogorov's strong law of large numbers, we have that

$$\frac{1}{n}\sum_{i=1}^{n} l_i - \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}_0(l_i) \to 0, \qquad \text{a.s..} \tag{4.5}$$

Notice that the above constant K is independent of θ and it can be shown that $(1/n) \cdot \sum_{i=1}^{n} \mathsf{E}_0(l_i(\theta))$ is equicontinuous in θ, then following the proof of Theorem 1 in Chiu et al. (1996) [4], it is easy to show the consistency of $\widehat{\theta}$.

The proof of (b) is essentially the same as that of Theorem 2 in Chiu et al. (1996) [4]. □

References

[1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis* (*Second Edition*), Wiley, New York, 1984.

[2] Barnard, J., McCulloch, R. and Meng, X.L., Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, *Statistica Sinica*, **10(4)**(2000), 1281–1312.

[3] Chen, Z. and Dunson, D.B., Random effects selection in linear mixed models, *Biometrics*, **59(4)** (2003), 762–769.

[4] Chiu, T.Y.M., Leonard, T. and Tsui, K.W., The matrix-logarithmic covariance model, *Journal of the American Statistical Association*, **91(433)**(1996), 198–210.

[5] Daniels, M.J. and Pourahmadi, M., Modeling covariance matrices via partial autocorrelations, *Journal of Multivariate Analysis*, **100(10)**(2009), 2352–2363.

[6] Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S.L., *Analysis of Longitudinal Data* (*Second Edition*), Oxford University Press, 2002.

[7] Joe, H., Generating random correlation matrices based on partial correlations, *Journal of Multivariate Analysis*, **97(10)**(2006), 2177–2189.

[8] Jennrich, R.I. and Schluchter, M.D., Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, **42(4)**(1986), 805–820.

[9] Kenward, M.G., A method for comparing profiles of repeated measurements, *Journal of the Royal Statistical Society: Series C* (*Applied Statistics*), **36(3)**(1987), 296–308.

[10] Leng, C., Zhang, W. and Pan, J., Semiparametric mean-covariance regression analysis for longitudinal data, *Journal of the American Statistical Association*, **105(489)**(2010), 181–193.

[11] Liang, K.Y. and Zeger, S.L., Longitudinal data analysis using generalized linear models, *Biometrika*, **73(1)**(1986), 13–22.

[12] Lin, T.I. and Wang, Y.J., A robust approach to joint modeling of mean and scale covariance for longitudinal data, *Journal of Statistical Planning and Inference*, **139(9)**(2009), 3013–3026.

[13] Molenberghs, G. and Kenward, M.G., *Missing Data in Clinical Studies* (*Wiley series: Statistics in Practice*), John Wiley and Sons, Inc., West Sussex, England, 2007.

[14] Pan, J. and Mackenzie, G., On modelling mean-covariance structures in longitudinal studies, *Biometrika*, **90(1)**(2003), 239–244.

[15] Potthoff, R.F. and Roy, S.N., A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51(3-4)**(1964), 313–326.

[16] Pourahmadi, M., Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, *Biometrika*, **86(3)**(1999), 677–690.

[17] Pourahmadi, M., Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix, *Biometrika*, **87(2)**(2000), 425–435.

[18] Rousseeuw, P.J. and Molenberghs, G., The shape of correlation matrices, *The American Statistician*, **48(4)**(1994), 276–279.

[19] Wang Y.G. and Carey, V., Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance, *Biometrika*, **90(1)**(2003), 29–41.

[20] Yang, R. and Berger, J.O., Estimation of a covariance matrix using the reference prior, *The Annals of Statistics*, **22(3)**(1994), 1195–1211.

[21] Ye, H. and Pan, J., Modelling of covariance structures in generalised estimating equations for longitudinal data, *Biometrika*, **93(4)**(2006), 927–941.

[22] Zhang, W. and Leng, C., A moving average Cholesky factor model in covariance modelling for longitudinal data, *Biometrika*, **99(1)**(2012), 141–150.

纵向数据中基于偏自相关的均值协方差同时建模

张伟平 刘玉婷 李瑞超

(中国科学技术大学统计与金融系, 合肥, 230026)

本文在纵向数据下提出一种均值–方差–相关矩阵的同时建模推断方法. 通过应用偏相关系数, 我们对相关系数矩阵进行无约束参数化, 并且能够自动保证估计的相关系数矩阵满足正定性. 在此基础上, 我们对参数提出了一种回归推断方法, 其具有简约性、可解释性和灵活性特点. 实际数据分析和模拟研究表明了所提方法是有效的.

关键词: 相关系数矩阵, 同时建模, 纵向数据分析, 偏相关系数.

学科分类号: O212.1.