Statistica Sinica: Supplement (SS-09-306)

ESTIMATION IN MULTIVARIATE t LINEAR MIXED MODELS FOR MULTIPLE LONGITUDINAL DATA

Wan-Lun Wang[1] and Tsai-Hung Fan[2]

[1]*Feng Chia University and* [2]*National Central University*

Supplementary Material

This note is a longer version of the printed paper. It contains some detailed derivations and a simulation study.

S1 The Score Vector and the Fisher Information Matrix

The score vector $\mathbf{s}_{\boldsymbol{\theta}}$ calculated by the first derivatives of ℓ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\omega}$ and ν, respectively, includes the following elements:

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\beta}} &= \sum_{i=1}^{N}\Big(\frac{\nu+n_i r}{\nu+\Delta_i}\Big)\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i, \\
[\mathbf{s}_{\boldsymbol{\omega}}]_l &= -\frac{1}{2}\sum_{i=1}^{N}\Big(\mathrm{tr}(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{il}) - \Big(\frac{\nu+n_i r}{\nu+\Delta_i}\Big)\boldsymbol{\epsilon}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{il}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i\Big), \\
s_{\nu} &= \frac{1}{2}\sum_{i=1}^{N}\Big(\mathcal{D}_{\mathcal{G}}(\frac{\nu+n_i r}{2}) - \mathcal{D}_{\mathcal{G}}(\frac{\nu}{2}) - \frac{n_i r}{\nu} - \log(1+\frac{\Delta_i}{\nu}) + \frac{(\nu+n_i r)\Delta_i}{(\nu+\Delta_i)\nu}\Big),
\end{aligned}$$

for $l=1,\ldots,g$, $g=q_2r(q_2r+1)/2+r(r+1)/2+p$, where

$$\dot{\boldsymbol{\Lambda}}_{il} = \frac{\partial\boldsymbol{\Lambda}_i(\mathbf{D},\boldsymbol{\Sigma},\boldsymbol{\phi})}{\partial\omega_l} = \begin{cases} \mathbf{Z}_i\frac{\partial\mathbf{D}}{\partial\omega_l}\mathbf{Z}_i^{\mathrm{T}} & \text{if } \omega_l=\mathrm{vech}(\mathbf{D}), \\ \frac{\partial\boldsymbol{\Sigma}}{\partial\omega_l}\otimes\mathbf{C}_i(\boldsymbol{\phi}) & \text{if } \omega_l=\mathrm{vech}(\boldsymbol{\Sigma}), \\ \boldsymbol{\Sigma}\otimes\dot{\mathbf{C}}_i(\boldsymbol{\phi}) & \text{if } \omega_l=\boldsymbol{\phi}, \end{cases}$$

and $\mathcal{D}_{\mathcal{G}}(\cdot)=d\log\Gamma(x)/dx$ is the digamma function. Note that $\partial\mathbf{D}/\partial\omega_l$ is 1 in the (l,s)th and (s,l)th elements of $\mathbf{D}$ as $\omega_l=d_{ls}$, and 0 otherwise; similarly for $\partial\boldsymbol{\Sigma}/\partial\omega_l$ when ω_l is the distinct element of $\boldsymbol{\Sigma}$; and $\dot{\mathbf{C}}_i(\boldsymbol{\phi})=-\mathbf{C}_i(\boldsymbol{\phi})[\partial\mathbf{C}_i^{-1}(\boldsymbol{\phi})/\partial\phi_v]\mathbf{C}_i(\boldsymbol{\phi})$ as $\omega_l=\phi_v$, $v=1,\ldots,p$, where $\partial\mathbf{C}_i^{-1}(\boldsymbol{\phi})/\partial\phi_v$ is given in (2.7) of Lin (2008).

S2 Proof of Theorem 1

Let $\boldsymbol{\epsilon}_i = \boldsymbol{\varepsilon}_i/\sqrt{\tau_i}$, where $\boldsymbol{\varepsilon}_i$ follows a $n_i r$-variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Lambda}_i$, and τ_i is independent of $\boldsymbol{\varepsilon}_i$ and follows $\nu^{-1}\chi^2_\nu$. Then, we have

$$(n_i r)^{-1}\Delta_i = (n_i r)^{-1}\boldsymbol{\epsilon}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\epsilon}_i = \frac{\boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\varepsilon}_i/(n_i r)}{\nu\tau_i/\nu},$$

which is the ratio of two independent chi-square variates, and thus $\Delta_i \sim n_i r\mathcal{F}(n_i r, \nu)$. By the fact that $1/\Delta_i \sim (n_i r)^{-1}\mathcal{F}(\nu, n_i r)$, we know that $\nu/(\nu+\Delta_i)$ can be written as the ratio of ν/Δ_i and $1+\nu/\Delta_i$ and thereby follows Beta$(\nu/2, n_i r/2)$ distribution.

The result in (i) follows directly from the property of Beta distribution, and the others are derived by taking expectations of desired quantities with respect to the density of $\mathbf{y}_i$. □

Consequently, the Fisher information matrix can be calculated as

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = E[-\partial^2\ell/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}] = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta\beta}} & \mathbf{J}_{\boldsymbol{\beta\alpha}} \\ \mathbf{J}^{\mathrm{T}}_{\boldsymbol{\beta\alpha}} & \mathbf{J}_{\boldsymbol{\alpha\alpha}} \end{bmatrix}, \qquad \text{(S2.1)}$$

where

$$\mathbf{J}_{\boldsymbol{\beta\beta}} = \sum_{i=1}^N \frac{(\nu+n_i r)}{(\nu+n_i r+2)}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\mathbf{X}_i,\ \mathbf{J}_{\boldsymbol{\beta\alpha}} = \mathbf{0}_{(q_1 r)\times(g+1)},\ \text{and } \mathbf{J}_{\boldsymbol{\alpha\alpha}} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\omega\omega}} & \mathbf{J}_{\boldsymbol{\omega}\nu} \\ \mathbf{J}^{\mathrm{T}}_{\boldsymbol{\omega}\nu} & J_{\nu\nu} \end{bmatrix}$$

has the following elements:

$$\begin{aligned}
[\mathbf{J}_{\boldsymbol{\omega\omega}}]_{ls} &= \frac{1}{2}\sum_{i=1}^N \frac{1}{\nu+n_i r+2}\Big((\nu+n_i r)\mathrm{tr}\big(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{il}\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{is}\big) - \mathrm{tr}\big(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{il}\big)\mathrm{tr}\big(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{is}\big)\Big),\\
[\mathbf{J}_{\boldsymbol{\omega}\nu}]_l &= -\sum_{i=1}^N \frac{1}{(\nu+n_i r)(\nu+n_i r+2)}\mathrm{tr}\big(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{il}\big),\\
J_{\nu\nu} &= \frac{1}{4}\sum_{i=1}^N \Big(\mathcal{T}_{\mathcal{G}}(\frac{\nu}{2}) - \mathcal{T}_{\mathcal{G}}(\frac{\nu+n_i r}{2}) - \frac{2n_i r(\nu+n_i r+4)}{\nu(\nu+n_i r)(\nu+n_i r+2)}\Big),
\end{aligned}$$

for $l, s = 1, \ldots, g$, with $\mathcal{T}_{\mathcal{G}}(x) = d^2\log\Gamma(x)/dx^2$ being a trigamma function.

S3 Proof of Theorem 3

(i) Based on the three-level hierarchical form (4.1), we utilize the Bayes' formula to compute the conditional density of $\mathbf{b}_i$ given $\mathbf{y}_i$ and τ_i. That is,

$$\begin{aligned} f(\mathbf{b}_i|\mathbf{y}_i,\tau_i) &\propto \exp\Big\{-\frac{\tau_i}{2}\Big[\mathbf{e}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{e}_i+\mathbf{b}_i^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{b}_i\Big]\Big\} \\ &\propto \exp\Big\{-\frac{\tau_i}{2}\Big[\mathbf{b}_i^{\mathrm{T}}(\mathbf{D}^{-1}+\mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{Z}_i)\mathbf{b}_i-2\mathbf{b}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\boldsymbol{\epsilon}_i\Big]\Big\} \\ &\propto \exp\Big\{-\frac{\tau_i}{2}\Big[\mathbf{b}_i-\mathbf{D}\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i\Big]^{\mathrm{T}}\Big(\mathbf{D}^{-1}+\mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{Z}_i\Big)\Big[\mathbf{b}_i-\mathbf{D}\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i\Big]\Big\}, \end{aligned}$$

which is a multivariate normal kernel with the desired mean vector and variance-covariance matrix.

(ii) Under the assumption of $\mathbf{y}_i|\tau_i \sim N_{n_ir}(\mathbf{X}_i\boldsymbol{\beta},\tau_i^{-1}\boldsymbol{\Lambda}_i)$ and $\tau_i \sim \text{Gamma}(\nu/2,\nu/2)$, we have

$$\begin{aligned} f(\tau_i|\mathbf{y}_i) &\propto f(\mathbf{y}_i|\tau_i)f(\tau_i) \\ &\propto \big|\tau_i^{-1}\boldsymbol{\Lambda}_i\big|^{-1/2}\exp\big\{-\frac{\tau_i}{2}\Delta_i\big\}\tau_i^{\nu/2-1}\exp\big\{-\frac{\nu}{2}\tau_i\big\} \\ &\propto \tau_i^{\frac{\nu+n_ir}{2}-1}\exp\big\{-\frac{\tau_i}{2}(\nu+\Delta_i)\big\}, \end{aligned}$$

which is the kernel of a gamma distribution with shape parameter $(\nu+n_ir)/2$ and inverse scale parameter $(\nu+\Delta_i)/2$. □

S4 Simulation Study

A small-scale simulation study was carried out to examine the finite sample properties of the proposed MtLMM. The main objective is to investigate the adequacy of the asymptotic results and to verify the ability of AIC and BIC in selecting the true model. In this study, we generate 100 Monte Carlo samples from a two-response model (3.1) with AR(1)-dependence errors, where $\mathbf{X}_i$ includes an intercept and scheduled visits of time (1 to 7), and $\mathbf{Z}_i$ contains an intercept only, namely $r = 2$, $q_1 = 2$, $q_2 = 1$ and $p = 1$. The presumed parameters are given as

$$\boldsymbol{\beta} = (1,2,-2,4)^{\mathrm{T}}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad \phi_1 = 0.5.$$

For the value of degrees of freedom (df), we take a small value ($\nu = 4$) to yield a heavy-tailed distribution.

The simulation was run with sample sizes $N = 20,\ 50$ and 100 to gauge the effect on the properties of the parameter estimates for an increasing number of observations. The three dependence structures considered in the study were UNC, AR(1) and AR(2).

For comparison purposes, each simulated data set was fitted six times under each combination of MLMM and MtLMM scenarios and three dependence structures. The detailed numerical results, including the average of the replicated ML estimates of all unknown parameters except for df and their average theoretical standard errors, are presented in Tables S1-S3. Notice that the median was chosen as an appropriate estimator for ν due to the strong skewness of the ML estimate. To evaluate the objective use of the criteria, the frequencies of each model preferred by AIC and BIC were also listed in the lower panel of tables.

As seen in these tables, the average estimated values have a tendency to reach their true values when the sample size is increasing. In particular, we can see a notable reduction of the standard errors for the fixed effects when using the MtLMM. We have also found that the standard errors of estimates will become smaller as sample size increases. The results do provide good asymptotic properties, at least for the set of parameters used in this study. As for model selection, there are (79%-87%) of AIC and (85%-97%) BIC agree with the specification of the true model. We may conclude that the BIC has a more consistent behavior in choosing the true model since the AIC has a tendency to pick models, which are over-parameterized. A similar indication applies to simulated data from more complicated models such as random effects with random intercepts and slopes. We skip to show all the details for making the paper concise.

Table S1. Simulation results based on 100 replications with sample size 20.

$\boldsymbol{\theta}$	True	MLMM						MtLMM					
		UNC		AR(1)		AR(2)		UNC		AR(1)		AR(2)	
		Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd
β_{10}	1	1.108	0.426	1.104	0.483	1.103	0.477	1.082	0.332	1.077	0.376	1.073	0.374
β_{11}	2	1.996	0.059	1.996	0.080	1.996	0.078	1.995	0.044	1.996	0.062	1.997	0.061
β_{20}	-2	-1.973	0.424	-1.985	0.478	-1.987	0.475	-1.997	0.326	-1.995	0.370	-1.995	0.368
β_{21}	4	3.990	0.059	3.994	0.080	3.994	0.079	3.999	0.044	4.001	0.062	4.001	0.061
d_{11}	1	2.369	0.849	1.912	0.962	1.872	0.981	1.317	0.536	1.006	0.546	1.009	0.586
d_{21}	0.25	0.599	0.620	0.302	0.667	0.335	0.677	0.372	0.358	0.214	0.367	0.215	0.383
d_{22}	1	2.324	0.834	1.746	0.918	1.772	0.967	1.248	0.509	0.929	0.517	0.930	0.557
σ_{11}	1	2.143	0.281	2.051	0.270	1.950	0.264	1.013	0.220	1.040	0.222	1.027	0.222
σ_{21}	0.5	1.007	0.217	0.992	0.209	0.978	0.206	0.505	0.139	0.521	0.141	0.516	0.141
σ_{22}	1	2.122	0.278	2.046	0.269	2.041	0.276	1.008	0.220	1.032	0.221	1.020	0.221
ϕ_1	0.5	—		0.482	0.084	0.479	0.083	—		0.489	0.088	0.486	0.088
ϕ_2	0	—		—		-0.035	0.088	—		—		-0.021	0.094
ν	4	—		—		—		4.200	1.963	4.325	2.122	4.308	2.134
Criterion		Frequency											
AIC		0		0		0		0		79		21	
BIC		0		1		0		0		85		14	

Table S2. Simulation results based on 100 replications with sample size 50.

$\boldsymbol{\theta}$	True	MLMM						MtLMM					
		UNC		AR(1)		AR(2)		UNC		AR(1)		AR(2)	
		Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd
β_{10}	1	1.062	0.285	1.057	0.316	1.058	0.315	1.059	0.212	1.064	0.237	1.064	0.237
β_{11}	2	1.994	0.037	1.996	0.051	1.996	0.051	1.994	0.028	1.993	0.039	1.993	0.039
β_{20}	-2	-2.020	0.281	-2.021	0.314	-2.020	0.313	-2.024	0.211	-2.020	0.237	-2.020	0.237
β_{21}	4	4.003	0.037	4.004	0.051	4.003	0.051	4.004	0.028	4.004	0.039	4.004	0.039
d_{11}	1	2.755	0.609	2.081	0.634	2.067	0.679	1.329	0.342	1.017	0.341	1.021	0.354
d_{21}	0.25	0.735	0.440	0.427	0.451	0.416	0.468	0.373	0.233	0.221	0.236	0.223	0.240
d_{22}	1	2.631	0.585	1.998	0.620	1.976	0.664	1.316	0.339	1.004	0.339	1.007	0.353
σ_{11}	1	2.003	0.166	1.957	0.163	1.938	0.166	0.982	0.137	0.998	0.138	0.994	0.139
σ_{21}	0.5	1.014	0.130	0.976	0.127	0.965	0.128	0.490	0.086	0.496	0.087	0.494	0.087
σ_{22}	1	2.010	0.167	1.968	0.164	1.948	0.167	0.987	0.138	1.010	0.139	1.005	0.140
ϕ_1	0.5	—		0.489	0.053	0.487	0.053	—		0.491	0.056	0.491	0.056
ϕ_2	0	—		—		-0.010	0.057	—		—		-0.009	0.060
ν	4	—		—		—		4.065	0.972	4.252	1.035	4.264	1.038
Criterion		Frequency											
AIC		0		0		0		0		85		15	
BIC		0		0		0		0		96		4	

Table S3. Simulation results based on 100 replications with sample size 100.

$\boldsymbol{\theta}$	True	MLMM						MtLMM					
		UNC		AR(1)		AR(2)		UNC		AR(1)		AR(2)	
		Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd	Est	Sd
β_{10}	1	0.998	0.199	1.000	0.223	1.000	0.223	0.994	0.149	0.992	0.168	0.992	0.168
β_{11}	2	2.005	0.027	2.005	0.037	2.005	0.037	2.005	0.020	2.006	0.028	2.006	0.028
β_{20}	-2	-2.025	0.199	-2.022	0.223	-2.021	0.222	-2.012	0.151	-2.010	0.170	-2.010	0.170
β_{21}	4	4.003	0.026	4.003	0.037	4.003	0.037	4.001	0.020	4.001	0.028	4.001	0.028
d_{11}	1	2.618	0.411	1.932	0.428	1.915	0.456	1.318	0.240	0.991	0.238	0.987	0.249
d_{21}	0.25	0.848	0.304	0.509	0.310	0.502	0.320	0.410	0.167	0.247	0.168	0.245	0.172
d_{22}	1	2.595	0.408	1.924	0.426	1.909	0.452	1.356	0.247	1.033	0.246	1.030	0.256
σ_{11}	1	1.994	0.117	1.945	0.114	1.937	0.117	0.969	0.096	0.994	0.097	0.992	0.098
σ_{21}	0.5	0.995	0.091	0.973	0.089	0.968	0.090	0.486	0.060	0.496	0.061	0.495	0.061
σ_{22}	1	1.965	0.115	1.936	0.114	1.924	0.116	0.981	0.097	1.007	0.098	1.005	0.099
ϕ_1	0.5	—		0.509	0.038	0.509	0.038	—		0.505	0.040	0.505	0.040
ϕ_2	0	—		—		-0.007	0.040	—		—		-0.003	0.043
ν	4	—		—		—		3.993	0.655	4.242	0.714	4.258	0.715
Criterion		Frequency											
AIC		0		0		0		0		87		13	
BIC		0		0		0		0		97		3	