Statistica Sinica: Supplement

DISCRETE LONGITUDINAL DATA MODELING WITH A MEAN-CORRELATION REGRESSION APPROACH

ChengYong Tang[1] and Weiping Zhang[2] and Chenlei Leng[3]

[1] *Temple Universityand* [2] *University of Science and Technology of China and* [3] *University of Warwick*

Supplementary Materials

The online supplementary materials contain the proofs of Theorem 1, 2 and 3 in the main paper, additional data analysis and simulations studies.

Computation of the score function. Note that the objective function is

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{1\le j<k\le m_i} l_{ijk}(\boldsymbol{\theta}),$$

where

$$l_{ijk}(\boldsymbol{\theta}) = log L_{ijk}(\boldsymbol{\theta}) = log \int_{z_{ij}^-}^{z_{ij}} \int_{z_{ik}^-}^{z_{ik}} \phi_2(\mathbf{u}; \rho_{ijk}) d\mathbf{u}$$

$$= log\Big(\Phi_2(z_{ij}, z_{ik}; \rho_{ijk}) - \Phi_2(z_{ij}^-, z_{ik}; \rho_{ijk}) - \Phi_2(z_{ij}, z_{ik}^-; \rho_{ijk}) + \Phi_2(z_{ij}^-, z_{ik}^-; \rho_{ijk})\Big)$$

and $\Phi_2(x, y; \rho)$ is the cdf of bivariate normal $N(0, 0, 1, 1, \rho)$, $z_{ij} = \Phi_1^{-1}\{F(y_{ij})\} = z_{ij}(\boldsymbol{\beta}, \psi)$, $z_{ij}^- = \Phi_1^{-1}\{F(y_{ij} - 1) = z_{ij}^-(\boldsymbol{\beta}, \psi)\}$, and denote $\boldsymbol{\eta} = (\boldsymbol{\beta}^{\mathrm{T}}, \psi)^{\mathrm{T}}$. We have

$$\begin{aligned}\frac{\partial l_{ijk}}{\partial \boldsymbol{\eta}} = \frac{1}{L_{ijk}} \frac{\partial L_{ijk}}{\partial \boldsymbol{\eta}} = \frac{1}{L_{ijk}} \Big(&\frac{\partial}{\partial \boldsymbol{\eta}} \Phi_2(z_{ij}, z_{ik}; \rho_{ijk}) - \frac{\partial}{\partial \boldsymbol{\eta}} \Phi_2(z_{ij}^-, z_{ik}; \rho_{ijk}) \\ &- \frac{\partial}{\partial \boldsymbol{\eta}} \Phi_2(z_{ij}, z_{ik}^-; \rho_{ijk}) + \frac{\partial}{\partial \boldsymbol{\eta}} \Phi_2(z_{ij}^-, z_{ik}^-; \rho_{ijk}) \Big).\end{aligned} \tag{A.1}$$

By the fact that

$$\begin{aligned}\frac{\partial \Phi_2(z_1, z_2; \rho)}{\partial \boldsymbol{\eta}} &= \frac{\partial \Phi_2(z_1, z_2; \rho)}{\partial z_1} \frac{\partial z_1}{\partial \boldsymbol{\eta}} + \frac{\partial \Phi_2(z_1, z_2; \rho)}{\partial z_2} \frac{\partial z_2}{\partial \boldsymbol{\eta}} \\ &= \phi(z_1) \Phi_1\Big(\frac{z_2 - \rho z_1}{\sqrt{1-\rho^2}}\Big) \frac{\partial z_1}{\partial \boldsymbol{\eta}} + \phi(z_2) \Phi_1\Big(\frac{z_1 - \rho z_2}{\sqrt{1-\rho^2}}\Big) \frac{\partial z_2}{\partial \boldsymbol{\eta}} \\ &= \Phi_1\Big(\frac{z_2 - \rho z_1}{\sqrt{1-\rho^2}}\Big) \frac{\partial F(y_1)}{\partial \boldsymbol{\eta}} + \Phi_1\Big(\frac{z_1 - \rho z_2}{\sqrt{1-\rho^2}}\Big) \frac{\partial F(y_2)}{\partial \boldsymbol{\eta}},\end{aligned} \tag{A.2}$$

where $z_i = \Phi_1^{-1}\{F(y_i)\}, i = 1, 2$, we can write out (A.1) easily.

Noting that for $j < k$, $\rho_{ijk} = \sum_{s=1}^{j} T_{ijs} T_{iks}$ and

$$\frac{\partial T_{its}}{\partial \boldsymbol{\gamma}} = \begin{cases} T_{its}[-tan(\omega_{its}) \frac{\partial \omega_{its}}{\partial \boldsymbol{\gamma}} + \sum_{l=1}^{s-1} \frac{1}{tan(\omega_{itl})} \frac{\partial \omega_{itl}}{\partial \boldsymbol{\gamma}}] & t > s > 1 \\ T_{its} \sum_{l=1}^{s-1} \frac{1}{tan(\omega_{itl})} \frac{\partial \omega_{itl}}{\partial \boldsymbol{\gamma}}, & t = s > 1 \\ -sin(\omega_{it1}) \frac{\partial \omega_{it1}}{\partial \boldsymbol{\gamma}}, & s = 1 \end{cases},$$

we can obtain the derivative of l_{ijk} with respect to $\boldsymbol{\gamma}$ as

$$\frac{\partial l_{ijk}}{\partial \boldsymbol{\gamma}} = \frac{1}{L_{ijk}}\frac{\partial L_{ijk}}{\partial \boldsymbol{\gamma}} = \frac{1}{L_{ijk}}\Big(\phi_2(z_{ij}, z_{ik}; \rho_{ijk}) - \phi_2(z_{ij}^-, z_{ik}; \rho_{ijk}) \\ - \phi_2(z_{ij}, z_{ik}^-; \rho_{ijk}) + \phi_2(z_{ij}^-, z_{ik}^-; \rho_{ijk})\Big)\frac{\partial \rho_{ijk}}{\partial \boldsymbol{\gamma}}. \tag{A.3}$$

Combining (A.1) and (A.3) leads to the score function $S_n(\boldsymbol{\theta})$.

The expected Hessian matrix. For the second derivatives of log-likelihood function, the formula is more complicated. However, it is easy to see that

$$\begin{aligned} E\mathbf{H}_n(\hat{\boldsymbol{\theta}}) &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{1\le j<k\le m_i} E\ddot{l}_{ijk}(\hat{\boldsymbol{\theta}}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{1\le j<k\le m_i} E\dot{l}_{ijk}(\hat{\boldsymbol{\theta}})\dot{l}_{ijk}^{\mathrm{T}}(\hat{\boldsymbol{\theta}}), \end{aligned} \tag{A.4}$$

thus $\mathbf{H}_n$ in (10) can be approximated by $\frac{1}{n}\sum_{i=1}^{n}\sum_{1\le j<k\le m_i}\dot{l}_{ijk}(\hat{\boldsymbol{\theta}})\dot{l}_{ijk}^{\mathrm{T}}(\hat{\boldsymbol{\theta}})$.

Proof of Theorem 1. The proof follows as a special case of the following proof for Theorem 2, and hence is omitted.

Proof of Theorem 2. Here we give a sketch of the proof. It is easy to see that $E_{\boldsymbol{\theta}}\mathbf{S}_n(\boldsymbol{\theta}) = 0$. Thus by Taylor expansion, we have

$$0 = \mathbf{S}_n(\hat{\boldsymbol{\theta}}) = \mathbf{S}_n(\boldsymbol{\theta}_0) + \dot{\mathbf{S}}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\dot{\mathbf{S}}_n = \partial \mathbf{S}_n^{\mathrm{T}}/\partial \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ is in a neighborhood of $\boldsymbol{\theta}_0$. Specially, we have $\tilde{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ when $n \to \infty$. Therefore, it is seen that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = [-\frac{1}{n}\dot{\mathbf{S}}_n(\tilde{\boldsymbol{\theta}})]^{-1}\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}_0).$$

From Central Limit Theorem, Assumption A1-A3, $E_{\boldsymbol{\theta}_0}\mathbf{S}_n(\boldsymbol{\theta}_0) = 0$ and the boundness of $Var_{\boldsymbol{\theta}_0}(\mathbf{S}_{ni}(\boldsymbol{\theta}_0)), i = 1, \ldots, n$, we have

$$\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}_0) \to N(0, \mathbf{J}(\boldsymbol{\theta}_0)).$$

By Assumption A3 and Slutsky's theorem, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal with asymptotic covariance matrix $\mathbf{G}(\boldsymbol{\theta}_0)$.

Proof of Theorem 3. Using a Taylor expansion of the log-pairwise likelihood function pl around $\boldsymbol{\theta}$, we obtain

$$pl(\hat{\boldsymbol{\theta}}) = pl(\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}\mathbf{S}_n(\boldsymbol{\theta}) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}(-n\mathbf{H}(\boldsymbol{\theta}))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1).$$

Notice that $0 = \mathbf{S}_n(\hat{\boldsymbol{\theta}}) = \mathbf{S}_n(\boldsymbol{\theta}) + (-n\mathbf{H}(\boldsymbol{\theta}))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(n^{1/2})$. We then have

$$pl(\hat{\boldsymbol{\theta}}) = pl(\boldsymbol{\theta}) + \frac{n}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}\mathbf{H}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1).$$

It can be rewritten via a partitioned matrix notation

$$
\begin{aligned}
pl(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) &= pl(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&\quad + \frac{n}{2}((\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)^{\mathrm{T}}, (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^{\mathrm{T}}) \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{pmatrix} + o_p(1).
\end{aligned}
\tag{A.5}
$$

Assuming that the null hypothesis is true, a Taylor expansion of the score $S_{n,2}$ around $(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2)$ gives

$$
0 = \mathbf{S}_{n,2}(\boldsymbol{\theta}_{1,0}, \tilde{\boldsymbol{\theta}}_2) = \mathbf{S}_{n,2}(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2) + (-n\mathbf{H}_{22})(\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(n^{1/2}).
$$

Equating this with the corresponding part of $S_n(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2)$, we find

$$
\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 = \mathbf{H}_{22}^{-1}\mathbf{H}_{21}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0}) + (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(n^{1/2}).
$$

Therefore under the null hypothesis, it is true that

$$
\begin{aligned}
2\{pl(\boldsymbol{\theta}_{1,0}, \tilde{\boldsymbol{\theta}}_2) - pl(\boldsymbol{\theta}_{1,0}, \tilde{\boldsymbol{\theta}}_2)\} &= n(\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^{\mathrm{T}} H_{22}(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2)(\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(1) \\
&= n[(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0})\mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0}) + 2(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0})^{\mathrm{T}}\mathbf{H}_{12}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \\
&\quad + (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)^{\mathrm{T}}\mathbf{H}_{22}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)] + o_p(1).
\end{aligned}
\tag{A.6}
$$

Combing (A.5) and (A.6) we have

$$
\begin{aligned}
2\{pl(\hat{\boldsymbol\theta}) - pl(\boldsymbol\theta_{1,0}, \tilde{\boldsymbol\theta}_2)\} &= 2\{pl(\hat{\boldsymbol\theta}) - pl(\boldsymbol\theta_{1,0}, \boldsymbol\theta_2)\} - 2\{pl(\boldsymbol\theta_{1,0}, \tilde{\boldsymbol\theta}_2) - pl(\boldsymbol\theta_{1,0}, \tilde{\boldsymbol\theta}_2)\} \\
&= n(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0})^{\mathrm{T}}(\mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21})(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0}) + o_p(1) \\
&= n(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0})^{\mathrm{T}}(\mathbf{H}^{11})^{-1}(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0}) + o_p(1).
\end{aligned}
$$

Because under the null hypothesis $\sqrt{n}(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0}) \to N(0, \mathbf{G}^{11})$, it follows from the properties of a multivariate normal distribution that

$$
n(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0})^{\mathrm{T}}(\mathbf{H}^{11})^{-1}(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{1,0}) \xrightarrow{d} \sum_{j=1}^{r} \lambda_j V_j,
$$

where $V_1, \ldots, V_r$ denote independent χ_1^2 random variables and $\lambda_1 \geq \cdots \geq \lambda_r$ are the eigenvalues of $(\mathbf{H}^{11})^{-1}\mathbf{G}^{11}$. The proof is completed.

Toenail data

We applied our mean-correlation regression method to analyze a data set from the toenail dermatophyte onychomycosis study (De Backer et al. (1996)). This data set consists of 294 participants in two treatment groups with a total of 1907 observations. Subjects were initially examined every month during a 12-week (3 months) treatment period, and then followed up further every 3 months for up to a total of 48 weeks (12 months). Due to various unknown reasons, in total there are 23.8% subjects dropping out, and consequently measurement numbers per subject range from 1 to 7. Therefore, this data set is unbalanced. The response variable of interest for our analysis is the severity of the infection of the toenail, coded as 0 (not severe) or 1 (severe). By analyzing this response variable, one aims to reveal the trend of the infection severity over time, and compare patterns, if any, between the two treatment groups. Following Molenberghs and Verbeke (2005), in the marginal model, we used the logistic model for the conditional mean function for the jth measurements of the ith subject:

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij}), \text{logit}(\pi_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij},$$

where T_i is the treatment indicator for subject i (1 for the experimental arm, 0 for the standard arm), t_{ij} is the time point at which the jth measurement is taken for the ith subject.

As for the correlation modeling, considering that the data set is unbalanced with homogeneously spaced time points for all subjects, we first investigated a reasonable model using a common 7×7 correlation matrix $\mathbf{R}$ by letting $\mathbf{R}_i = \mathbf{R}$ for all subjects. Thus the equivalent unknown parameters for $\mathbf{R}$ by the parametrization (2.4) were ω_{jk} $(1 \leq j < k \leq 7)$.

Then the pairwise likelihood approach was applied to obtain estimators $\tilde{\omega}_{jk}$, leading to an estimated correlation matrix. The plot of the function $\tan(\pi/2-\tilde{\omega}_{jk})$ versus the time lag was given in Figure 5 (a) with solid dots, suggesting some monotone decreasing associations. Clearly, this method for incorporating the correlations involves $7 \times 6/2 = 21$ parameters.

Now let us demonstrate the application of the parsimonious correlation regression. Suggested by Figure 5 (a) and the composite likelihood versions of Bayesian information criterion (BIC) described by Gao and Song (2010), we link these angles with covariates via the parsimonious model specified in (5) using a quadratic polynomial function of the time lag between measurements with unknown parameters $\gamma_0, \gamma_1, \gamma_2$. The estimated parameters of the mean-correlation joint model with estimated standard deviation shown in the subscript were $\hat{\beta}_0 = -0.5565_{0.1711}, \hat{\beta}_1 = 0.0236_{0.2407}, \hat{\beta}_2 = -0.1830_{0.0232}, \hat{\beta}_3 = -0.0774_{0.0344}$, suggesting that the time was a significant covariate in the mean model, while the evidence for the treatment effect and its interaction with time was not statistically significant. For comparisons, we also obtained a GEE estimates of the parameters in the same mean model with unstructured working correlations: $\tilde{\beta}_0 = -0.6898_{0.1679}$, $\tilde{\beta}_1 = 0.0828_{0.2430}$, $\tilde{\beta}_2 = -0.1483_{0.0283}$ and $\tilde{\beta}_3 = -0.1043_{0.0514}$. We found that the two sets of estimates are largely comparable with each other. The estimated parameters in the correlation regression model were $\hat{\gamma}_0 = 3.0236_{0.2750}, \hat{\gamma}_1 = -0.4690_{0.0658}, \hat{\gamma}_2 = 0.0204_{0.0043}$, all highly significant. Denoted by $\hat{\omega}_{jk}$ the estimated angles from the parsimonious model, Figure 5 (a) also shows the plot of the fitted angles $\tan(\pi/2 - \hat{\omega}_{jk})$ versus time lag, which indicates a competent fitting of the angles with far fewer parameters where only 3 parameters are involved compared with 21 parameters in a common correlation matrix $\mathbf{R}$. Figure 5 (b) indicates, not surprisingly, that the correlation decreases as the time lag increases, suggesting a high correlation between the severity of the infection at current visit with the those at the nearest visit times.

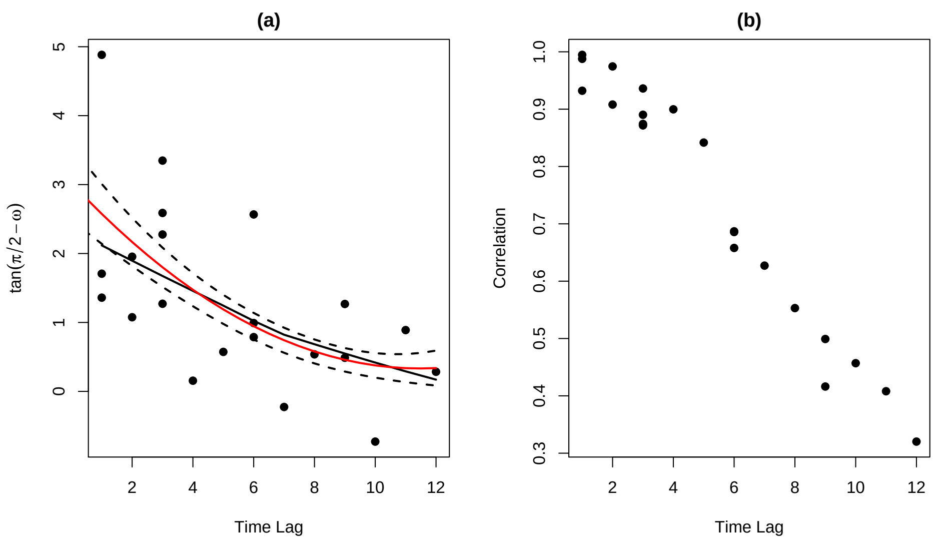

Figure 5: The toenail data: (a) plot of fitted angles $\tan(\pi/2 - \hat{\omega}_{jk})$ versus time lag, (b) plot of fitted correlations versus time lag. In panel (a), solid dots are fitted angles with a common correlation matrix for all subjects with parametrization (4), the solid black line is from fitting a LOWESS curve to the solid dots, the solid red line is from the proposed model, and the dashed curves represent asymptotic 95% confidence intervals.

Additional simulations

Study 4. We generated n observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$, each with dimension m_i set as the two cases in Study 1. In this study we considered a Gaussian copula model in which the marginal distributions $F_{ij}(j = 1, ..., m)$ are negative binomial as $y_{ij} \sim NegBin(\delta, \mu_{ij})$ with mean μ_{ij} and variance $\mu_{ij} + \mu_{ij}^2/\delta$, where $\delta > 0$ is the over-dispersion parameter. The mean was parameterised as $\mu_{ij} = \exp(\mathbf{x}_{ij}^{\mathrm{T}}\beta)$ to allow dependence on covariates, and the variance exceeds its mean (i.e. overdispersion). The covariate x_{ij1} and x_{ij2} were bivariate normal with correlation 0.5. The angles in the correlation matrix was set as $\omega_{ijk} = \pi/2 - \text{atan}(\gamma_0 + w_{ijk1}\gamma_1)$ with $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2\}^{\mathrm{T}}$. The true parameters were taken as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) = (1, -0.5, 0.5)$, $\delta = 4$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2) = (0.5, -0.3, 0.5)$.

Table 4 shows that all the biases for the proposed method are small and that the SD and SE are quite close, especially for large n. Interestingly, the MLEs perform slightly better in this study for Case I, but we observed that it took much more time to obtain them. For Case II, the large bias of the MLEs suggest again that the MLE may encounter severe numerical problems when the multi-dimensional integrations are computed. In terms of the estimation efficiency of the parameters in the mean model, the proposed PLEs again performs very competitively compared with the GEE method with unstructured correlations in this case.

Table 4: Simulation results for Study 4. Mean bias (MB) and standard deviation (SD) of each parameter us reported. SE is the average standard error calculated using the formula in Theorem 2. PL: Partial Likelihood; FL: Full Likelihood; GEE: Generalized Estimating Equation.

	Pairwise Likelihood			Full Likelihood			GEE		
n	50	100	200	50	100	200	50	100	200
				Case I					
MB_{β_0}	-0.002	-0.002	-0.001	-0.004	-0.004	-0.004	-0.008	-0.002	-0.001
SD	(0.047)	(0.058)	(0.046)	(0.051)	(0.063)	(0.050)	(0.093)	(0.058)	(0.044)
SE	0.056	0.041	0.029	-	-	-	-	-	-
Std	(0.006)	(0.005)	(0.003)	-	-	-	-	-	-
MB_{β_1}	0.008	-0.004	-0.003	0.004	-0.006	-0.004	-0.001	-0.001	-0.001
SD	(0.018)	(0.031)	(0.022)	(0.024)	(0.032)	(0.024)	(0.044)	(0.027)	(0.021)
SE	0.032	0.023	0.016	-	-	-	-	-	-

Std	(0.005)	(0.003)	(0.002)	-	-	-	-	-	-
MB_{β_2}	0.003	-0.002	-0.002	0.001	-0.004	-0.004	-0.002	0.001	-0.000
SD	(0.010)	(0.035)	(0.025)	(0.019)	(0.036)	(0.025)	(0.044)	(0.030)	(0.020)
SE	0.032	0.023	0.016	-	-	-	-	-	-
Std	(0.005)	(0.003)	(0.002)	-	-	-	-	-	-
MB_{δ}	0.561	0.313	0.111	0.282	0.395	0.224	1.407	0.791	0.500
SD	(0.746)	(1.048)	(0.640)	(0.453)	(1.138)	(0.750)	(2.559)	(1.848)	(1.416)
SE	1.128	0.724	0.469	-	-	-	-	-	-
Std	(0.382)	(0.273)	(0.103)	-	-	-	-	-	-
MB_{γ_0}	-0.006	-0.004	-0.001	0.003	-0.091	-0.093	-	-	-
SD	(0.113)	(0.079)	(0.058)	(0.025)	(0.187)	(0.193)	-	-	-
SE	0.100	0.073	0.051	-	-	-	-	-	-
Std	(0.013)	(0.007)	(0.003)	-	-	-	-	-	-
MB_{γ_1}	-0.019	-0.009	0.011	0.002	0.447	0.4318	-	-	-
SD	(0.654)	(0.433)	(0.332)	(0.081)	(0.459)	(0.426)	-	-	-
SE	0.471	0.332	0.231	-	-	-	-	-	-
Std	(0.083)	(0.049)	(0.022)	-	-	-	-	-	-
MB_{γ_2}	0.052	0.022	-0.006	0.004	-0.413	-0.387	-	-	-
SD	(0.764)	(0.522)	(0.398)	(0.112)	(0.401)	(0.329)	-	-	-
SE	0.549	0.384	0.266	-	-	-	-	-	-
Std	(0.107)	(0.066)	(0.031)	-	-	-	-	-	-
					Case II				
MB_{β_0}	-0.009	-0.001	-0.005	-0.010	0.001	-0.004	-0.012	0.000	-0.004
SD	(0.090)	(0.068)	(0.047)	(0.088)	(0.065)	(0.045)	(0.088)	(0.065)	(0.045)
SE	0.031	0.020	0.014	-	-	-	-	-	-
Std	(0.060)	(0.027)	(0.013)	-	-	-	-	-	-
MB_{β_1}	0.000	-0.001	-0.001	0.000	0.000	-0.001	-0.001	0.000	-0.001
SD	(0.046)	(0.032)	(0.023)	(0.044)	(0.031)	(0.022)	(0.046)	(0.031)	(0.023)
SE	0.037	0.026	0.018	-	-	-	-	-	-
Std	(0.007)	(0.004)	(0.002)	-	-	-	-	-	-
MB_{β_2}	-0.001	-0.001	0.000	-0.000	-0.000	0.000	-0.001	-0.000	0.000
SD	(0.049)	(0.032)	(0.024)	(0.046)	(0.030)	(0.022)	(0.047)	(0.031)	(0.024)
SE	0.037	0.026	0.019	-	-	-	-	-	-
Std	(0.007)	(0.004)	(0.002)	-	-	-	-	-	-
MB_{δ}	0.777	0.310	0.120	0.864	0.405	0.228	1.340	0.0.907	0.600
SD	(1.953)	(1.051)	(0.659)	(2.073)	(1.176)	(0.831)	(2.624)	(1.984)	(1.461)
SE	1.358	0.770	0.502	-	-	-	-	-	-

Std	(0.944)	(0.306)	(0.123)	-	-	-	-	-	-
MB_{γ_0}	-0.001	-0.011	-0.005	-0.063	-0.070	-0.057	-	-	-
SD	(0.145)	(0.067)	(0.070)	(0.114)	(0.056)	(0.069)	-	-	-
SE	0.125	0.086	0.061	-	-	-	-	-	-
Std	(0.019)	(0.008)	(0.005)	-	-	-	-	-	-
MB_{γ_1}	-0.042	-0.044	0.021	0.407	0.458	0.379	-	-	-
SD	(0.826)	(0.096)	(0.382)	(0.550)	(0.086)	(0.364)	-	-	-
SE	0.591	0.403	0.285	-	-	-	-	-	-
Std	(0.118)	(0.061)	(0.032)	-	-	-	-	-	-
MB_{γ_2}	0.070	-0.037	-0.018	-0.477	-0.520	-0.442	-	-	-
SD	(0.994)	(0.657)	(0.451)	(0.643)	(0.480)	(0.418)	-	-	-
SE	0.704	0.475	0.335	-	-	-	-	-	-
Std	(0.165)	(0.081)	(0.043)	-	-	-	-	-	-

References

De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day a double-blind comparative trial. *Br. J. Dermatol.* **134**, 16–7.

Gao, X. and Song, P.X.K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *J. Amer. Statist. Assoc.* **105**, 1531-1540.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag.