Robust linear mixed models using the skew $t$ distribution with application to schizophrenia data

Tsung I. Lin

(Joint work with Hsiu J. Ho)

Department of Applied Mathematics and Institute of Statistics
National Chung Hsing University, Taiwan

Dec. 1, 2010
1 Background knowledge

2 Motivating Example: the Schizophrenia Data

3 The skew $t$ linear mixed model (STLMM)
   - Multivariate skew-normal and skew-$t$ distributions
   - A generalization of the SNLMM (Lin and Lee, 2008)

4 ML Estimation via the AECM Algorithm

5 Prediction of Random Effects

6 Application: The Schizophrenia Example Revisited

7 Conclusion
Skewed to the Right vs. Left

Skewed to the right

Skewed to the Left
OLD FAITHFUL GEYSER DATA (LIN ET AL., 2007)

Faithful data

probability

NORMIX
SNMIX

T. I. Lin (NCHU)
Definition (Azzalini, 1985; Scan. J.) A random variable $Y$ follows a univariate skew normal distribution with location $\xi$, scale variance $\sigma^2$ and skewness $\lambda \in \mathbb{R}$ if $Y$ has the following density function:

$$f(y \mid \xi, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \xi}{\sigma}\right) \Phi\left(\lambda \frac{y - \xi}{\sigma}\right),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and the cdf of $N(0, 1)$, respectively.

- Denoted by $Y \sim SN(\xi, \sigma^2, \lambda)$.
- If $\lambda = 0$, $Y \sim N(\xi, \sigma^2)$.
- If $Z \sim SN(0, 1, \lambda) \equiv SN(\lambda)$, then $Z^2 \sim \chi_1^2$. 

T.I. Lin (NCHU)
STANDARD SKEW NORMAL DENSITIES

Densities of standard Skew-Normal with various skewnesses

\[ f(x) \]

\[ \lambda = -3 \]
\[ \lambda = -2 \]
\[ \lambda = -1 \]
\[ \lambda = 0 \]
\[ \lambda = 1 \]
\[ \lambda = 2 \]
\[ \lambda = 3 \]
Univariate skew normal distribution (SN)

\[ f(x|\lambda) = 2 \phi(x) \Phi(\lambda x) \]

\[ \delta = 0 \]

\[ \lambda = \delta(1 - \delta^2)^{-1/2} \]
**THE UNIVARIATE SKEW t DISTRIBUTION**

- **Stochastic representation**

  \[
  Y = \xi + \sigma \frac{Z}{\sqrt{\tau}}, \quad Z \sim SN(\lambda), \quad \tau \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \equiv \frac{\chi_{\nu}^2}{\nu}, \quad Z \perp \tau.
  \]

- \[ Y | \tau \sim SN\left(0, \frac{1}{\tau}, \lambda\right) \]

A random variable \( Y \) is said to follow a univariate skew \( t \) distribution with location parameter \( \xi \), scale parameter \( \sigma^2 \), skewness parameter \( \lambda \in \mathbb{R} \) and degrees of freedom \( \nu \) if \( Y \) has the following density function

\[
f(y) = \int f(y | \tau)f(\tau)d\tau = \frac{2}{\sigma} t_{\nu}(\eta) T_{\nu+1} \left( \lambda \eta \sqrt{\frac{\nu + 1}{\eta^2 + \nu}} \right), \quad \eta = \frac{y - \xi}{\sigma},
\]

where \( t_{\nu}(\cdot) \) and \( T_{\nu}(\cdot) \) denote the pdf and the cdf of the Student’s \( t \) distribution with degrees of freedom \( \nu \), respectively.
(Normal) Linear mixed models ((N)LMM; Laid and Ware, 1982)

\[ Y = X\beta + Zb + \varepsilon \]
\[ b \sim \text{Normal} \quad \varepsilon \sim \text{Normal} \]

Some robust extensions

1. *t* linear mixed model (TLMM; Pinheiro et al., 2001)
2. \ldots
3. Skew normal linear mixed model (SNLMM; Lin and Lee, 2008)
4. \ldots
5. Skew *t* linear mixed model (STLMM; Today’s talk)
This study involves a double-blind clinical trial with randomization among four treatments for 245 patients with acute schizophrenia.

- three doses (low, medium and high) of a new therapy (NT)
- a standard therapy (ST)

The data were collected from 13 clinical centers.

Response variable: the Brief Psychiatric Rating Scale (BPRS) at baseline (week zero), and at weeks 1, 2, 3, 4 and 6 of treatment.

BPRS score: $0 \rightarrow 108$ (severe).

We present here only the comparison between the 57 patients on high dose of NT and the 61 patients on ST.
Figure 1: Trajectories of schizophrenia levels for the data. The thicker solid line indicates the mean profile in the treatment.
Motivating Example: the Schizophrenia Data

Preliminary Analysis

LMM with curvilinear-trend fixed effects and normal-distributed random effects and within-subject errors

\[ y_{ij} = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 NT_i + b_{0i} + b_{1i} t_j + \varepsilon_{ij}; \]
\[ i = 1, \ldots, 118, \ j = 1, \ldots, 6, \]  

where

- \( y_{ij} = \text{BPRS}/10 \) at the \( j \)th time point for the \( i \)th subject;
- \( t_j \) is taken as \((\text{time} - 3)/10\) with time being measured in week from the baseline;
- \( NT_i \) an indicator variable of NT for subject \( i \);
- \( \beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top \) is the fixed effects of explanatory variables;
- \( b_i = (b_{0i}, b_{1i})^\top \) is the random effects vector for the \( i \)th subject; and
- \( \varepsilon_{ij} \) is the within-subject error.
Figure 2: Histograms and corresponding normal quantile plots of the empirical Bayes estimates of random effects obtained from fitting LMM to the schizophrenia data.
**Figure 3:** Residuals versus fitted values (upper panels) and the normal quantile plots corresponding to residuals (lower panels).
Model and Method

Multivariate Skew Normal (MSN) Distribution

- Skew normal distribution, \( Z \sim SN_p(\mu, \Sigma, \lambda) \), has

The density

\[
f(Z) = 2 \phi_p(Z \mid \mu, \Sigma) \Phi\left(\lambda^\top \Sigma^{-1/2}(Z - \mu)\right).
\]

The stochastic representation

\[
Z = \mu + \Sigma^{1/2} \delta \gamma + \Sigma^{1/2} (I_p - \delta \delta^\top)^{1/2} U, \quad \gamma \perp U
\]

where \( \delta = \lambda / \sqrt{1 + \lambda^\top \lambda} \), \( \gamma \sim T \mathcal{N}(0, 1; (0, \infty)) \equiv |\mathcal{N}(0, 1)| \), \( U \sim \mathcal{N}_p(0, I_p) \) and the symbol ‘\( \perp \)’ indicates independence.

\( \lambda = 3 \quad \rho = -0.9 \)
\[ \lambda = 0.5 \quad \rho = 0 \]
Multivariate Skew $t$ (MST) Distribution

- The MST distribution, $Y \sim St_p(\mu, \Sigma, \lambda, \nu)$, can be represented by

The stochastic representation of skew $t$ distribution

$$Y = \mu + \frac{1}{\sqrt{\tau}} Z, \quad Z \perp \tau$$

where $Z \sim SN_p(0, \Sigma, \lambda)$ and $\tau \sim \text{Gamma}(\nu/2, \nu/2)$.

- $Y | \tau \sim SN_p(\mu, \Sigma/\tau, \lambda)$
- Integrating $\tau$ from the joint density of $(Y, \tau)$ yields

The marginal density of $Y$

$$f(Y | \mu, \Sigma, \lambda; \nu) = 2 \cdot t_p(Y | \mu, \Sigma; \nu) \cdot T\left(\lambda^\top \Sigma^{-1/2}(Y - \mu) \sqrt{\frac{\nu + \rho}{\nu + \Delta}}; \nu + \rho\right)$$

(2)

where $\Delta = (Y - \mu)^\top \Sigma^{-1} (Y - \mu)$. 

I. Lin (NCHU)
National Chung Hsing University
Dec. 1, 2010

15 / 35
\( \nu = 3, \ \lambda = 0.5 \)
$\lambda = (3, 3), \ \nu = 2$
The Australian Institute of Sport (AIS) data (Lin (2010, Statist. Comput.))

Model Multivariate skew-normal and skew-$t$ distributions

MVNMIX

MSTMIX
Applications of Flow cytometry

MVNMIX

MSTMIX

T. I. Lin (NCHU)
National Chung Hsing University
Skew $t$ Linear Mixed Models

**STLMM**

The model considered here can be written as

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i$$

along with the assumption of

$$\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} \sim St_{q+n_i} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \Gamma & 0 \\ 0 & C_i \end{bmatrix}, \begin{bmatrix} \lambda \\ 0 \end{bmatrix}, \nu \right),$$

$i = 1, \ldots, N$.

- $\Gamma$ is a $q \times q$ unstructured positive definite matrix, $C_i = C_i(\rho)$ is an $n_i \times n_i$ dependence matrix, a function of small set of parameters $\rho = (\rho_1, \ldots, \rho_g)$ and depends on $i$ only through its dimension $n_i$. 
The hierarchical formulation I of STLMM (3)

\[ Y_i \mid (\gamma_i, \tau_i) \sim \mathcal{N}_{n_i} \left( X_i \beta + \gamma_i d_i, \frac{\sigma^2}{\tau_i} \psi_i \right) \]

\[ \gamma_i \mid \tau_i \sim \mathcal{T} \mathcal{N} \left( 0, \frac{\sigma^2}{\tau_i}; (0, \infty) \right) \]

\[ \tau_i \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \] (5)

The hierarchical formulation II of STLMM (3)

\[ Y_i \mid (b_i, \gamma_i, \tau_i) \sim \mathcal{N}_{n_i} \left( X_i \beta + Z_i b_i, \frac{\sigma^2}{\tau_i} C_i \right) \]

\[ b_i \mid (\gamma_i, \tau_i) \sim \mathcal{N}_q \left( \xi \gamma_i, \frac{\sigma^2}{\tau_i} V \right) \]

\[ \gamma_i \mid \tau_i \sim \mathcal{T} \mathcal{N} \left( 0, \frac{\sigma^2}{\tau_i}; (0, \infty) \right) \]

\[ \tau_i \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \] (6)
The distribution of $Y_i$

$$Y_i \sim St_{n_i}(X_i\beta, \sigma^2 \Lambda_i, \alpha_i, \nu),$$

where

$$\alpha_i = \frac{\Lambda_i^{-1/2} d_i}{\sqrt{1 - d_i^\top \Lambda_i^{-1} d_i}}.$$

The mean and covariance matrix of $Y_i$

$$E(Y_i) = X_i\beta + \sigma \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} d_i$$

$$\text{cov}(Y_i) = \sigma^2 \left( \frac{\nu}{\nu - 2} \Lambda_i - \frac{\nu}{\pi} \left( \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right)^2 d_i d_i^\top \right)$$
Applying Bayes’ rule to (6) yields

\[
\mathbf{b}_i \mid (\gamma_i, \tau_i, \mathbf{Y}_i) \sim \mathcal{N}_q\left(\mathbf{u}_i \gamma_i + \mathbf{v}_i, \frac{\sigma^2}{\tau_i} \mathbf{\Sigma}_i\right)
\]

\[
\gamma_i \mid (\tau_i, \mathbf{Y}_i) \sim \mathcal{TN}\left(\kappa_i \mathbf{A}_i, \frac{\sigma^2}{\tau_i} \kappa_i^2; (0, \infty)\right)
\]

\[
f(\tau_i \mid \mathbf{Y}_i) = \frac{\Phi(\sigma^{-1} \sqrt{\tau_i} \mathbf{A}_i)}{\mathcal{T}(c_{0i}; \nu + n_i)} \frac{g\left(\tau_i \left| \frac{\nu + n_i}{2}, \frac{\nu + \sigma^{-2} \Delta_i}{2}\right\right)}{g\left(\tau_i \left| \frac{n_i}{2}, \frac{\nu}{2}\right\right)}
\]  

The conditional expectation of latent variables in (7) are very useful for the alternating expectation-conditional maximization (AECM) algorithm (Meng and van Dyk, 1997).
A diagram of the AECM Algorithm

Estimation

\[
AECM
\]

\[
\ell_c^{[1]}(\theta_1 \mid Y_{\text{aug}}^{[1]})
\]

\[
\ell(\theta \mid Y)
\]

\[
\ell_c^{[2]}(\theta_2 \mid Y_{\text{aug}}^{[2]})
\]

\[
\hat{\theta}^{(k+1)}
\]

\[
\hat{\theta}^{(k+\frac{1}{2})}
\]

\[
Q[1](\theta_1 \mid \hat{\theta}^{(k)})
\]

\[
Q[2](\theta_2 \mid \hat{\theta}^{(k+\frac{1}{2})})
\]

\[
\text{Cycle 1}
\]

\[
\text{Cycle 2}
\]

\[
\text{stopping rule}
\]

\[
\text{T.I. Lin (NCHU)}
\]

National Chung Hsing University

Dec. 1, 2010

22 / 35
The Asymptotic Covariance Matrix

The approximate variance-covariance matrix of the ML estimates can be evaluated by inverting the expected information or the observed information matrix.

The observed information matrix, \( I_o = -\partial^2 \sum_{i=1}^{N} \log f(Y_i|\theta) / \partial \theta \partial \theta^T \), can be obtained by using Louis’ (1982) formula

\[
I_o(\theta | Y) = \sum_{i=1}^{N} E[I_{ci}|Y_i] - \sum_{i=1}^{N} \text{cov}[s_{ci}|Y_i]
\]  

(8)

where \( s_{ci} \) and \( I_{ci} \) are the individual score vector and negative Hessian matrix with respect to the complete data log-likelihood function formed from the single observation \( Y_i \), respectively.

\[
SE(\hat{\theta}_i) \approx \sqrt{[\text{diag}(I_o^{-1}(\hat{\theta} | Y))]_{ii}}.
\]
Prediction of random effects

- We consider an empirical Bayes-based approach to the prediction of random effects that is useful for examining subject-specific quantities of interest.

- The minimum mean-squared error (MSE) predictor of $b_i$, obtained by the conditional mean of $b_i$ given $Y_i$, is

$$\hat{b}_i(\theta) = E[b_i|Y_i] = \kappa_i A_i \left( 1 + \frac{1}{c_{-2,i}} \frac{t(c_{-2,i}; \nu + n_i - 2)}{T(c_{0i}; \nu + n_i)} \right) u_i + v_i$$

(9)

where $A_i$, $\kappa_i$, $c_{ji}$, $u_i$ and $v_i$ are defined in (7).

- The empirical Bayes estimates of $b_i$, $\hat{b}_i$, can be obtained by substituting the ML estimate $\hat{\theta}$ into (9). As a consequence, it leads to

$$\hat{b}_i = \hat{b}_i(\hat{\theta}).$$
Prediction of Missing Values

- In longitudinal studies, missing values arise frequently due partly to early withdrawal or failure to meet scheduled appointments.
- The resulting missingness yields an unbalanced pattern with unequal number of measurements or intermittent missing values for each subject.
- Under the MAR mechanism, we provide a conditional predictor for imputing intermittent missing values in model (1).
- We partition $Y_i (n_i \times 1)$ into two components $(Y^o_i, Y^m_i)$, where $Y^o_i (n^o_i \times 1)$ and $Y^m_i ((n_i - n^o_i) \times 1)$ denote the observed and missing components, respectively.
- To facilitate computation, two auxiliary permutation matrices are introduced such that $Y^o_i = O_i Y_i$ and $Y^m_i = M_i Y_i$, where $O_i (n^o_i \times n_i)$ and $M_i ((n_i - n^o_i) \times n_i)$ can be extracted from an $n_i$-dimensional identity matrix $I_{n_i}$ corresponding to row positions of $Y^o_i$ and $Y^m_i$ in $Y_i$. 
From (6), we have that

\[ Y_i^m \mid (Y_i^o, b_i, \gamma_i, \tau_i) \sim \mathcal{N}_{n_i-n_i^o}(\mu_{i^m-o}, \frac{\sigma^2}{\tau_i} C_{i^m-o}^o) \]

\[ Y_i^o \mid (b_i, \gamma_i, \tau_i) \sim \mathcal{N}_{n_i^o}(O_i \mu_i, \frac{\sigma^2}{\tau_i} O_i C_i O_i^\top) \]

where \( \mu_i = X_i \beta + Z_i b_i \), \( S_i^o = O_i^\top (O_i C_i O_i^\top)^{-1} O_i \),
\[ \mu_{i^m-o} = M_i (\mu_i + C_i S_i^o (Y_i - \mu_i)) \) and \( C_{i^m-o} = M_i (C_i - C_i S_i^o C_i) M_i^\top \).

The minimum MSE predictor of \( Y_i^m \), \( \hat{Y}_i^m(\theta) \), can be expressed as

\[ \hat{Y}_i^m(\theta) = E[Y_i^m \mid Y_i^o] = M_i (X_i \beta + C_i S_i^o (Y_i - X_i \beta) + W_i E[b_i \mid Y_i^o]) \] (10)

The prediction of \( Y_i^m \), \( \hat{Y}_i^m \), is then obtained by substituting ML estimates \( \hat{\theta} \) into (10), leading to

\[ \hat{Y}_i^m = \hat{Y}_i^m(\hat{\theta}). \]
Application: The Schizophrenia Example Revisited

- Based on the preliminary analysis, we are motivated to advocate the use of STLMM as a promising tool to analyze this data set.

- We compare the ML results under the STLMM with those obtained under the reduced LMM, TLMM and SNLMM models.

- In the STLMM setting, we modify model (1) with the random effects $b_i = (b_{0i}, b_{1i})^\top$ and error terms $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^\top$ jointly distributed as

$$
\begin{bmatrix}
  b_i \\
  \varepsilon_i
\end{bmatrix}
\overset{\text{ind}}{\sim}
\text{St}_{2+n_i}
\left(
\begin{bmatrix}
  0 \\
  0
\end{bmatrix},
\sigma^2
\begin{bmatrix}
  \Gamma & 0 \\
  0 & I_{n_i}
\end{bmatrix},
\begin{bmatrix}
  \lambda \\
  0
\end{bmatrix},
\nu
\right),
$$

where $i = 1, \ldots, 118$. 
Table 1: ML estimation results for four competitive models with the associated standard errors in parentheses, where $F$, with distinct elements $F_{ij}$, is the square root of $\Gamma$ such that $\Gamma = F^2$, and $\delta = \lambda / \sqrt{1 + \lambda \top \lambda}$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LMM</th>
<th>TLMM</th>
<th>SNLMM</th>
<th>STLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>Sd</td>
<td>Est</td>
<td>Sd</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>2.4759</td>
<td>0.2097</td>
<td>2.2174</td>
<td>0.1830</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1.6404</td>
<td>0.3828</td>
<td>-1.9038</td>
<td>0.3420</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>6.9039</td>
<td>0.6454</td>
<td>6.3670</td>
<td>0.5162</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.1413</td>
<td>0.2108</td>
<td>-0.0517</td>
<td>0.1809</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.6197</td>
<td>0.0159</td>
<td>0.4682</td>
<td>0.0278</td>
</tr>
<tr>
<td>$F_{11}$</td>
<td>1.9056</td>
<td>0.1642</td>
<td>2.2499</td>
<td>0.1943</td>
</tr>
<tr>
<td>$F_{12}$</td>
<td>1.0391</td>
<td>0.1969</td>
<td>1.3670</td>
<td>0.2287</td>
</tr>
<tr>
<td>$F_{22}$</td>
<td>4.1474</td>
<td>0.6242</td>
<td>4.7102</td>
<td>0.6154</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\nu$</td>
<td>—</td>
<td>4.7755</td>
<td>1.1455</td>
<td>—</td>
</tr>
<tr>
<td>$m$</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>$\ell(\hat{\Theta})$</td>
<td>-793.83</td>
<td>-759.27</td>
<td>-777.87</td>
<td>-737.30</td>
</tr>
<tr>
<td>AIC</td>
<td>1603.65</td>
<td>1536.55</td>
<td>1575.74</td>
<td>1496.60</td>
</tr>
<tr>
<td>BIC</td>
<td>1625.82</td>
<td>1561.48</td>
<td>1603.44</td>
<td>1527.09</td>
</tr>
</tbody>
</table>
Figure 4: Scatter plot of estimated random effects superimposed on a set of contour lines, together with two summary histograms of their marginal densities.
To detect outlying observations, the predictive distribution of $\tau_i$ can be used as concise indicators for detecting outliers with prior expectation of 1.

We regard the $i$th participant as an outlier in the population if the 95% prediction interval $(q_{0.025}, q_{0.975})$ for $\tau_i$ does not cover 1.

From the distribution $f(\tau_i \mid Y_i)$ in (7), the empirical quantile $q_p$ ($0 \leq p \leq 1$) is defined by the relation

$$
\int_{0}^{q_p} \frac{\Phi(\hat{\sigma}^{-1}\sqrt{\tau_i}\hat{A}_i)}{T(\hat{c}_0; \nu + n_i)} g\left(\tau_i \left| \frac{\hat{\nu} + n_i}{2}, \frac{\hat{\nu} + \hat{\sigma}^{-2}\hat{\Delta}_i}{2}\right.\right) d\tau_i = p
$$

We compute the 95% prediction interval $(q_{0.025}, q_{0.975})$ for each $\tau_i$ and found that there are 16 intervals does not cover 1. We conclude that they (the 16 patients) are suspect outliers.
Figure 5: Trajectories of schizophrenia levels for the data. The thicker solid line indicates the mean profile in the treatment. The red dashed lines indicates suspect outlying observations.
Diagnostics

- For LMM and SNLMM, a formal measure for checking the distributional assumption is through Mahalanobis-like distance

\[ \Delta^*_i = \sigma^{-2} \hat{e}_i^\top \hat{\Lambda}_i^{-1} \hat{e}_i, \]

where \( \hat{e}_i = Y_i - X_i \hat{\beta} \), which has an asymptotic chi-square distribution with \( n_i \) df. Checking these two models can be achieved by constructing a chi-square (Healy’s) plot.

- To assess the fitness of TLMM and STLMM, it can be shown that \( \Delta^*_i / n_i \) follows a \( F \) distribution with \( n_i \) and \( \nu \) df.

- Thus, one can construct another Healy-type plot (or the F plot) by plotting the ordered \( F \) statistics against the quantiles of \( F(n_i, \nu) \) distribution for nominal values \( (i - 0.5)/N, i = 1, \ldots, N \).

- One can examine whether the corresponding Healy’s plot resembles a straight line through the origin having unit slope (identity line).
Figure 6: Healy’s plot for assessing the goodness-of-fit of fitted models.
Conclusion

- We propose a robust approach to LMM based on the MST distribution, called **STLMM**, as a powerfully tool to handle longitudinal data with asymmetric and discrepant behaviors in repeated measurements.

- We have described **two flexible hierarchical representations** for STLMM and presented a computationally efficient **AECM algorithm** for carrying out ML estimation.

- The **empirical Bayes** estimation procedure for the prediction of random effects is easy to implement once the ML estimates are obtained.

- Numerical results show that the proposed model is **overwhelmingly suited** to the illustrated schizophrenia example.