THE MIXED EFFECTS TREND VECTOR MODEL

Mark de Rooij
Leiden University
Psychological Institute
Methodology and Statistics Group

CARMÉ 2011 - Rennes, France
Mixed effects approaches to longitudinal data

1. Mixed effects models explicitly model individual change across time

2. No need to have balanced design or equally spaced measurements
   - Individuals may vary in their number of measurements by design or due to attrition
   - Individuals with missing responses can be included under a missing at random assumption

3. Straightforward to allow for between individual variation in the timing of measurements

4. Flexible in the relationship between time and response (polynomial functions)

5. Can allow for clustering at higher levels (repeated measurements of children in classrooms)

6. There exist generalizations for non-normal data (generalized linear mixed models).
Longitudinal multinomial data

1. Longitudinal multinomial data are often gathered in the social sciences.
   - In consumer science, for example, consumers are often asked for their preferred type of soup (brand) which may be one of a long list.
   - In criminology, interest is often in the type of crimes that people commit and not just in whether a crime is committed.
   - In political science interest is often in vote transitions between political parties which may be numerous.

   These are just a few examples where the number of categories of the response variable may be large.

2. We would like to model these data with a mixed effects model such that we have a mechanism for the dependency among the responses. The subject specific parameters are assumed to be random effects from a Normal distribution.

3. The multinomial distribution for a response variable with $C$ categories can be considered as a multivariate binomial distribution, with dimensionality $C - 1$. 
Some notation

The sample consists of \(n\) subjects and for each subject \(i\) there are measurements on \(n_i\) occasions. Let \(G_{it}\) denote the \(t\)-th observation for subject \(i\), with \(G_{it} = c\) \((c = 1, \ldots, C)\) and response probabilities \(\pi_{itc} = P(G_{it} = c)\).

Furthermore let \(g_{it}\) be the corresponding vector \(g_{it} = [g_{it1}, \ldots, g_{itC}]^T\) with \(g_{itc} = 1\) if subject \(i\) \((i = 1, \ldots, n)\) at time point \(t\) \((t = 1, \ldots, n_i)\) chooses category \(c\) \((c = 1, \ldots, C)\), zero otherwise.

We have two design vectors

- \(x_{it}\) is the design vector for the fixed effects;
- \(z_{it}\) is the design vector for the random effects.

The conditional distribution of \(g_{it}\) given a set of subject specific parameters \(u_i\), \(f(g_{it}|u_i)\), is the multinomial distribution, which belongs to the multivariate exponential family, with expectation

\[
E(g_{it}|u_i) = \pi_{it} = [\pi_{it1}, \ldots, \pi_{itC}]^T.
\]
The mixed effects multinomial baseline category logit model

The probabilities are related to a linear predictor by the vector of link functions $h_l(\cdot)$, i.e.

$$\pi_{it} = h_l(\eta_{it}),$$

and $h_l(\cdot) = [h_{l1}(\cdot), \ldots, h_{lC}(\cdot)]$, where $h_{lc}(\cdot)$ is

$$h_{lc}(\eta_{it1}, \ldots, \eta_{itC}) = \frac{\exp(\eta_{itc})}{\sum_{h} \exp(\eta_{ith})}.$$

The $c$-th linear predictor is given by

$$\eta_{itc} = \alpha_c + x_{it}^T \beta_c + z_{it}^T u_{ic},$$

where $x_{it}$ is the design vector for the fixed effects, $z_{it}$ is the design vector for the random effects, and $\alpha_c, \beta_c$ are fixed effect parameters. In order to identify the model, one set of parameters is fixed to zero, i.e. $\alpha_1 = 0$, $\beta_1 = 0$, and $u_{i1} = 0$. A multivariate normal distribution is assumed for the random effects, i.e.

$$u_{ic} \sim N(0, \Sigma), c = 2, \ldots, C.$$
Housing condition across time by group: proportions and sample size

<table>
<thead>
<tr>
<th>Group</th>
<th>Status</th>
<th>Baseline</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Street</td>
<td>.555</td>
<td>.186</td>
<td>.089</td>
<td>.124</td>
</tr>
<tr>
<td></td>
<td>Community</td>
<td>.339</td>
<td>.578</td>
<td>.582</td>
<td>.455</td>
</tr>
<tr>
<td></td>
<td>Independent</td>
<td>.106</td>
<td>.236</td>
<td>.329</td>
<td>.421</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>180</td>
<td>161</td>
<td>146</td>
<td>145</td>
</tr>
<tr>
<td>Incentive</td>
<td>Street</td>
<td>.442</td>
<td>.093</td>
<td>.121</td>
<td>.120</td>
</tr>
<tr>
<td></td>
<td>Community</td>
<td>.414</td>
<td>.280</td>
<td>.146</td>
<td>.228</td>
</tr>
<tr>
<td></td>
<td>Independent</td>
<td>.144</td>
<td>.627</td>
<td>.732</td>
<td>.652</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>181</td>
<td>161</td>
<td>157</td>
<td>158</td>
</tr>
</tbody>
</table>
Solution for the Mixed effects MBCL model

For the MHRP data we fitted a model with quadratic time trend and random intercepts. The linear predictor equals

$$\eta_{itc} = \alpha_c + G_i \beta_1 + T_{it} \beta_2 + T_{it}^2 \beta_3 + G_i T_{it} \beta_4 + G_i T_{it}^2 \beta_5 + u_{ic},$$

where $G_i$ is an indicator for group membership ($G_i = 1$ for incentive) for participant $i$, and $T_{it}$ represents the time variable.

Parameter estimates are:

<table>
<thead>
<tr>
<th>Effect</th>
<th>C/S</th>
<th>SE</th>
<th>I/S</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.5960</td>
<td>0.2223</td>
<td>-2.5836</td>
<td>0.3657</td>
</tr>
<tr>
<td>Time</td>
<td>0.4565</td>
<td>0.0579</td>
<td>0.5571</td>
<td>0.0708</td>
</tr>
<tr>
<td>Time Squared</td>
<td>-0.0147</td>
<td>0.0023</td>
<td>-0.0159</td>
<td>0.0027</td>
</tr>
<tr>
<td>Incentive</td>
<td>0.7054</td>
<td>0.3150</td>
<td>1.0882</td>
<td>0.4649</td>
</tr>
<tr>
<td>Incentive × Time</td>
<td>-0.2450</td>
<td>0.0802</td>
<td>0.1569</td>
<td>0.0949</td>
</tr>
<tr>
<td>Incentive × Time squared</td>
<td>0.0079</td>
<td>0.0033</td>
<td>-0.0069</td>
<td>0.0037</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.5448</td>
<td>0.2002</td>
<td>2.3149</td>
<td>0.2241</td>
</tr>
</tbody>
</table>

The correlation between the two random intercepts equals 0.696.
Problems with the Mixed effects MBCL model

1. These models may become computational very intensive when there are two or more random effects, and computational infeasible when there are more than five or six random effects.

2. These models rely on the untestable assumption that random coefficients come from a multivariate normal distribution. Results may be biased when this assumption is violated.

3. It is not at all straightforward to interpret the parameters associated with the random effects.

4. The interpretation of regression coefficients is not simple, especially in cases with interactions and/or higher order treatment of variables. The interpretation is further complicated because the coefficients refer to contrasts of categories of the response variable with a baseline category.
The mixed effect trend vector model

The probabilities are related to squared distances by the vector of link functions $h(\cdot)$, i.e.

$$\pi_{it} = h(\delta_{it}),$$

with

$$\delta_{it} = [\delta_{it1}, \ldots, \delta_{itC}]^T,$$

and $h(\cdot) = [h_1(\cdot), \ldots, h_C(\cdot)]$, where $h_c(\cdot)$ is the Gaussian decay function

$$h_c(\delta_{it1}, \ldots, \delta_{itC}) = \frac{\exp(-\delta_{itc})}{\sum_l \exp(-\delta_{ilt})}.$$
The mixed effect trend vector model

Let us now define the \( m \)-th linear predictor

\[
\eta_{itm} = \alpha_m + x_{it}^T \beta_m + z_{it}^T u_{im},
\]

which in multidimensional scaling terms gives the ideal point for subject \( i \) at time point \( t \) on dimension \( m \). We assume a multivariate normal distribution for the random effects of dimension \( m \), i.e.

\[
u_{im} \sim N(0, \Sigma_m)
\]

and we assume that the random effects for dimension \( m \) are uncorrelated with those of dimension \( m' \) (\( m \neq m' \)). For random intercept models this is without loss of information, since the axis can always be rotated to principal axis.

Finally, define category points \( \gamma_{cm} \) and the squared Euclidean distance between ideal points and category points links to the transformed expected values, i.e.

\[
\delta_{itc} = \sum_{m=1}^{M} (\eta_{itm} - \gamma_{cm})^2.
\]

The (mixed effect) trend vector model equals the (mixed effect) MBCL model when \( M = C' - 1 \).
Estimation

It is assumed that conditional on the random effects the responses are independent.

To obtain maximum likelihood estimates of the model parameters $\beta_{jm}$, $\gamma_{cm}$, and $\Sigma_m$ we use marginal maximum likelihood estimation

$$L = \prod_i \int \cdots \int f(g_i|u_i; \beta_m, \gamma_m) f(u_i; \Sigma) du_i.$$

This likelihood can be approximated using Gauss-Hermite quadrature, where the integral is replaced by a weighted summation over a set of nodes. The more nodes are used the better the approximation, but the slower the algorithm. The approximated likelihood is maximized using a quasi-Newton algorithm.

Prediction of the random effects can be done using expected a posteriori estimation.
Graphical display of MBCL solution

Incentive group
Control group

CARME 2011 - Rennes, France
The analysis of asymmetry with explanatory variables


<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDA</td>
<td>PvdA</td>
<td>VVD</td>
<td>GL</td>
<td>SP</td>
<td>D66</td>
<td>CU</td>
</tr>
<tr>
<td>2003</td>
<td>CDA</td>
<td>365</td>
<td>18</td>
<td>31</td>
<td>2</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>PvdA</td>
<td>15</td>
<td>309</td>
<td>9</td>
<td>12</td>
<td>111</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>VVD</td>
<td>76</td>
<td>8</td>
<td>186</td>
<td>1</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>GL</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>46</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>SP</td>
<td>6</td>
<td>14</td>
<td>0</td>
<td>9</td>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>D66</td>
<td>7</td>
<td>14</td>
<td>16</td>
<td>11</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>CU</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>474</td>
<td>374</td>
<td>243</td>
<td>81</td>
<td>290</td>
<td>35</td>
<td>72</td>
</tr>
</tbody>
</table>

For each of the 1569 participants we do not only have information on the two choices but also measurements on six background variables.
The analysis of asymmetry with explanatory variables

- **Income**: Some people think that the differences in incomes in our country should be increased. Others think that they should be decreased. Where would you place yourself on a line from 1 to 7, where 1 means differences in income should be increased and 7 means that differences in income should be decreased?

- **Asylum**: Some people think that the Netherlands should allow more asylum seekers to enter. Others that the Netherlands should send asylum seekers who are already staying here back to their country of origin. Where would you place yourself on a line from 1 to 7, where 1 means there are more asylum seekers allowed to enter and 7 means asylum seekers are send back?

- **Crime**: People think differently about the way the government fights crime. Where would you place yourself on a line from 1 to 7, where at the beginning of the line the parties are that think the government is acting too tough on crime and at the end of the line the parties are that think the government should be tougher on crime?

- **Nuclear**: Some people think that nuclear power plants are the solution to a shortage of energy in the future. Others think nuclear power plants shouldn’t be build, because the dangers are too great. Where would you place yourself on a line from 1 to 7, where 1 means nuclear power plants should be build quickly an 7 means that they shouldn’t be build?
The analysis of asymmetry with explanatory variables

- *Foreign*: In the Netherlands some think that foreigners should be able to live in the Netherlands while preserving their own culture. Others think that they should fully adapt to Dutch culture. Where would you place yourself on a line from 1 to 7, where 1 means preservation of own culture for foreigners and 7 means that they should fully adapt?

- *Europe*: Some people and parties think that the European unification should go further. Others think that the European unification has already gone too far. Where would you place yourself on a line from 1 to 7, where 1 means that the European unification should go even further and 7 that the unification has already gone too far?

Model:

\[ \eta_{itm} = \alpha_m + u_{0i.m} + T_{it} \beta_{1m} + I_{i} \beta_{2m} + A_{i} \beta_{3m} + C_{i} \beta_{4m} + N_{i} \beta_{5m} + F_{i} \beta_{6m} + E_{i} \beta_{7m} \]
The analysis of asymmetry with explanatory variables

Regression weights and test statistics for the explanatory variables.

<table>
<thead>
<tr>
<th>Effect</th>
<th>dim</th>
<th>Estimate</th>
<th>SE</th>
<th>LRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.4508</td>
<td>0.083</td>
<td>29.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0644</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>1</td>
<td>0.7541</td>
<td>0.084</td>
<td>206.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.2787</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>Asylum</td>
<td>1</td>
<td>-0.4656</td>
<td>0.077</td>
<td>52.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0310</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>Crime</td>
<td>1</td>
<td>-0.1490</td>
<td>0.072</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.1300</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>Nuclear</td>
<td>1</td>
<td>0.3679</td>
<td>0.052</td>
<td>73.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0261</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>1</td>
<td>-0.2525</td>
<td>0.065</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.1120</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>1</td>
<td>0.2612</td>
<td>0.056</td>
<td>28.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0244</td>
<td>0.049</td>
<td></td>
</tr>
</tbody>
</table>
The analysis of asymmetry with explanatory variables
We proposed mixed effect trend vector models for longitudinal categorical data.

In maximum dimensionality the trend vector model equals the MBCL model, but provides a graphical display of the results.

The trend vector model has the possibility of dimension reduction.

The integral dimension is also reduced, which makes ML fitting using quadrature methods feasible.

Neat graphical displays are obtained, which are readily interpretable.

The model can be applied to a wide range of problems with dichotomous, ordered and unordered categorical responses.

For ordered response variables the models could be fitted in one dimension, possibly with order restrictions on the class-point coordinates. For such data the ordinality can be tested.

The model assumes independence given the random effects, further research on the examination of this assumption is needed.