Small Area Estimation under the Growth Curve model

Innocent Ngaruye,

Linköping University, Sweden
Outline

- Introduction
- The model formulation
- Estimation of model parameters
- Prediction of random effects
- Simulation study example
- Further research
- Some references
The term *Growth Curve Modeling* has been used in different contexts to refer to a wide array of statistical models for repeated measures data.

It has long played a significant role in empirical research within the developmental sciences, particularly in studying between-individual differences and within-individual patterns of change over time.
We propose to apply this model in SAE settings to get a model which borrows strength across both small areas and over time by incorporating simultaneously the effects of areas and time interaction.

This model accounts for repeated surveys, group individuals and random effects variation. The estimation is discussed with a likelihood based approach and a simulation study is conducted.
We consider repeated measurements on variable of interest $y$ for $p$ time points, $t_1, ..., t_p$ from the finite population $U$ of size $N$ partitioned into $m$ disjoint subpopulations or domains $U_1, ..., U_m$ called small areas of sizes $N_i$, $i = 1, ..., m$ such that $\sum_{i=1}^{m} N_i = N$.

We also assume that in every area, there are $k$ different groups of units of size $N_{ig}$ for group $g$ such that $\sum_{i}^{m} \sum_{g=1}^{k} N_{ig} = N$.

We draw a sample of size $n$ in all small areas such that the sample of size $n_i$ is observed in area $i$ and $\sum_{i}^{m} \sum_{g=1}^{k} n_{ig} = n$ and we suppose that we have auxiliary data $x_{ij}$ of $r$ variables (covariates) available for each population unit $j$ in all $m$ small areas.
The model formulation (cont’d)

- The model at Small Area level is given by

\[
Y_i = A B_i C_i + 1 \gamma' X_i + 1 u'_i + E_i,
\]

\[
u_i \sim N_{N_i}(0, \sigma^2 u I),
\]

\[
E_i \sim N_{p, N_i}(0, \sigma^2 e I, I_{N_i}),
\]

where \( A \) and \( C_i \) are respectively within-individual and between-individual design matrices for fixed effects given by

\[
A = \begin{pmatrix}
1 & t_1 & \cdots & t_1^{q-1} \\
1 & t_2 & \cdots & t_2^{q-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & t_p & \cdots & t_p^{q-1}
\end{pmatrix},
\]

\[
C_i = \begin{pmatrix}
1 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1
\end{pmatrix}
\]
The corresponding model at population level for all small areas can be expressed as

\[
Y = \begin{pmatrix} A & B & C \end{pmatrix} + \begin{pmatrix} 1 \gamma' \left[ \mathbf{I}_r : \mathbf{I}_r : \cdots : \mathbf{I}_r \right] \end{pmatrix} \begin{pmatrix} X \end{pmatrix} + \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} u' \end{pmatrix} + \begin{pmatrix} E \end{pmatrix}
\]

or

\[
Y = ABC + 1 \gamma' DX + 1u' + E,
\]

for \( D = \left[ \mathbf{I}_r : \mathbf{I}_r : \cdots : \mathbf{I}_r \right] \)
In order to transform (2) to a model which is easier to estimate, we transform the design matrix $\mathbf{A}$ into a new matrix $\mathbf{A}_1$ with two parts $\mathbf{A}_1 = [\mathbf{1} : \mathbf{H}]$ and the parameter matrix into a new matrix $\Xi = [\xi_1 : \Xi_2]$ conformably such that

$$
\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{1}) \oplus \mathcal{C}(\mathbf{H}) \text{ with } \mathcal{C}(\mathbf{H}) = \mathcal{C}(\mathbf{1})^\perp \cap \mathcal{C}(\mathbf{A})
$$

One way of this transformation is given below

$$
\begin{pmatrix}
1 & t_1 & \cdots & t_{q-1} \\
1 & t_2 & \cdots & t_{q-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & t_p & \cdots & t_{q-1}
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & t_1 - \bar{t} & \cdots & t_{q-1} - \bar{t} \\
1 & t_2 - \bar{t} & \cdots & t_{q-1} - \bar{t} \\
\vdots & \vdots & \ddots & \vdots \\
1 & t_p - \bar{t} & \cdots & t_{q-1} - \bar{t}
\end{pmatrix}
$$

Innocent Ngaruye, Linköping University, Sweden

Small Area Estimation under the Growth Curve model
We come up with the model

\[ Y = 1\xi' C + H\Xi_2 C + 1\gamma' DX + 1u' + E \]

and make a one-to-one transformation

\[
\begin{pmatrix}
1'Y \\
H'Y \\
A^o'Y
\end{pmatrix}
= \begin{pmatrix}
p\xi' C + p\gamma' DX + pu' + 1'E \\
H'H\Xi_2 C + H'E \\
A^o'E
\end{pmatrix},
\]

where \( A^o \) for a matrix \( A \) is such that \( A^o'A = 0 \) and \( \mathcal{C}(A^o) = \mathcal{C}(A)^\perp \).
After calculation, the maximum likelihood estimators are given by

\[ \hat{\Xi}_2 = (H'H)^{-1} H'YC' (CC')^{-1} + (H'H)^{\circ} T_1 + H'HT_2 (CC')^{\circ'} \]
\[ \hat{\gamma}' = \frac{1}{p} \left[ 1'YX'D' - 1'YC'(CC')^{-1}CX'D' - pT_3 (CC')^{\circ} CX'D' \right] \]
\[ \times \left[ DXX'D' - DXC'(CC')^{-1}C \right]^{-1} \]
\[ \hat{\xi}_1' = \left( \frac{1}{p} 1'Y - \hat{\gamma}'DX \right) C'(CC')^{-1} + T (CC')^{\circ} \]

for some matrices \( T, T_1, T_2 \) and \( T_3 \) of proper sizes.
Once $\hat{\xi}_1'$ and $\hat{\Xi}_2$ are obtained, we can then find the parameter matrix $B$ by solving the linear system

$$1\hat{\xi}_1'C + H\hat{\Xi}_2C = A\hat{B}C.$$ 

Since, the matrices $A$ and $C$ are of full rank, then

$$\hat{B} = (A'A)^{-1}A'(1\hat{\xi}_1'C + H\hat{\Xi}_2C)C'(CC')^{-1}.$$
Given the covariance structure of $\mathbf{Y}$

$$\mathbf{\Sigma} = \mathbf{1}\mathbf{\Sigma}_u\mathbf{1}' + \mathbf{\Sigma}_e = m\sigma_u^2 \mathbf{1}\mathbf{1}' + \sigma_e^2 \mathbf{I}_p,$$

and its inverse

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma_e^2} \left( \mathbf{I}_p - \frac{m\sigma_u^2}{mp\sigma_u^2 + \sigma_e^2} \mathbf{1}\mathbf{1}' \right).$$

We find the maximum likelihood estimator of the variance component expressed by

$$\hat{\sigma}_u^2 = \frac{\text{tr}\{\mathbf{1}\mathbf{1}'\mathbf{W}\} - Np\sigma_e^2}{Nmp^2},$$

where

$$\mathbf{W} = (\mathbf{Y} - \mathbf{ABC} - \mathbf{1}\gamma'\mathbf{D}\mathbf{X})(\mathbf{Y} - \mathbf{ABC} - \mathbf{1}\gamma'\mathbf{D}\mathbf{X})'.$$
Under the theory of linear model and normal distribution, the best linear predictor of $u$ that minimizes the mean square error is the conditional mean $E[u|Y]$ given by

$$E[u|Y] = E[u] + \text{Cov}(u', Y) \text{Cov}^{-1}(Y)(Y - E[Y]).$$

Thus,

$$\hat{u} = \hat{\sigma}_u^2 1' \Sigma^{-1} (Y - \hat{A}\hat{B}\hat{C} - 1\hat{\gamma}'D'X)$$

$$= \frac{\hat{\sigma}_u^2}{mp\hat{\sigma}_u^2 + \sigma_e^2} 1'(Y - \hat{A}\hat{B}\hat{C} - 1\hat{\gamma}'D'X)$$
We consider 6 small areas and draw a sample with the following sample sizes.

<table>
<thead>
<tr>
<th>Area</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_{11}=52$</td>
<td>$n_{12}=48$</td>
<td>$n_1=100$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21}=60$</td>
<td>$n_{22}=60$</td>
<td>$n_2=120$</td>
</tr>
<tr>
<td>3</td>
<td>$n_{31}=30$</td>
<td>$n_{32}=40$</td>
<td>$n_3=70$</td>
</tr>
<tr>
<td>4</td>
<td>$n_{41}=46$</td>
<td>$n_{42}=22$</td>
<td>$n_4=68$</td>
</tr>
<tr>
<td>5</td>
<td>$n_{51}=65$</td>
<td>$n_{52}=65$</td>
<td>$n_5=130$</td>
</tr>
<tr>
<td>6</td>
<td>$n_{61}=50$</td>
<td>$n_{62}=62$</td>
<td>$n_6=112$</td>
</tr>
<tr>
<td>m=6</td>
<td>$g_1=303$</td>
<td>$g_2=297$</td>
<td>n=600</td>
</tr>
</tbody>
</table>

We assume $p = 4$ and $r = 3$. 
The design matrices are

\[ A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \rightarrow H = \begin{pmatrix} -1.5 \\ -0.5 \\ 0.5 \\ 1.5 \end{pmatrix}, \]

\[ C = \begin{pmatrix} C_1 & \cdots & 0 \\ 0 & \cdots & C_6 \end{pmatrix} \quad \text{for} \quad C_i = \begin{pmatrix} 1'_{n_{i1}} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : 1'_{n_{i2}} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}, \quad i = 1, \ldots, 6; \]
The parameter matrices are

\[ \xi_1' = (20 \ 21 \ 22 \ 23 \ 24 \ 25 \ 26 \ 27 \ 28 \ 29 \ 30 \ 31), \]
\[ \Xi_2 = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12), \]

\[ B = A^{-1} \left( 1 \xi_1'C + H \Xi_2 C \right) C^{-1} \]

\[ = \begin{pmatrix}
17.5 & 16 & 14.5 & 13 & 11.5 & 10 & 8.5 & 7 & 5.5 & 4 & 2.5 & 1 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12
\end{pmatrix}, \]

and

\[ \gamma = \begin{pmatrix}
1 \\
2 \\
3
\end{pmatrix}, \sigma_u^2 = 5, \quad \sigma_e^2 = 6. \]
Then, the data are generated from

\[ Y \sim N_{p,n}(ABC + 1\gamma'DX, \Sigma, I_n), \]

where the matrix of covariates \( X \) is generated with random elements. The following MLEs are obtained:


\[ \hat{\Xi}_2 = (1.1151, 2.0824, 3.0320, 3.6376, 4.6384, 5.7882, 7.0238, 7.8776, 9.0386, 10.1256, 10.8561, 11.9422) \]
Simulation study Example (cont’d)


\[ \hat{\sigma}^2_u = 5.0061, \quad \hat{\gamma} = \begin{pmatrix} 1.0093 \\ 1.9501 \\ 3.0469 \end{pmatrix}, \quad \hat{A}\hat{B}\hat{C} = \begin{pmatrix} 18.5 & \cdots & 18 & \cdots & 13 \\ 19.5 & \cdots & 20 & \cdots & 25 \\ 20.5 & \cdots & 22 & \cdots & 37 \\ 21.5 & \cdots & 24 & \cdots & 49 \end{pmatrix} \]
Further research

- After obtaining all unknown parameters, then we can find directly the target small area characteristics of interest such as the small area totals and small area means.
- In further research, we want to test the efficiency, the distribution and all properties of the estimators.
- We wish also to study the possible time correlation.
Some references


Some references


Note: This work is being jointly conducted in the framework of my PhD research together with my supervisors:

Dietrich von Rosen and Martin Singull to whom I owe my acknowledgements.

THANKS !!!!!!!!