Biometrical modeling of twin and family data

Sophia Rabe-Hesketh
Graduate School of Education & Graduate Group in Biostatistics
University of California, Berkeley
Institute of Education, University of London

joint work with Anders Skrondal and Håkon Gjessing

German Stata Users Group Meeting
Berlin, June 2010
Outline

- Genetic variance components model: ACDE
- Liability model for binary traits
- Models for twin designs
  - Assumptions and two parameterizations (P1, P2) as mixed/multilevel models
  - Continuous adult height: P1 ACE, P2 ACE
  - Continuous neuroticism: P2 ADE
  - Binary hay-fever status: P2 ADE & AE
- Models for nuclear family designs
  - Continuous birth weight data
**Genetic variance components models: ACDE**

- $y_{ij}$ is continuous trait or phenotype for member $i$ of family $j$
  \[ y_{ij} = x'_{ij} \beta + A_{ij} + D_{ij} + C_{ij} + \epsilon_{ij} \]

- **Error components**
  - $A_{ij} \sim N(0, \sigma_A^2)$: Additive genetic, potentially correlated
  - $D_{ij} \sim N(0, \sigma_D^2)$: Dominance genetic, potentially correlated
  - $C_{ij} \sim N(0, \sigma_C^2)$: Common environment, potentially correlated
  - $\epsilon_{ij} \sim N(0, \sigma_E^2)$: Unique environment, independent
  - $A_{ij}, D_{ij}, C_{ij}, \epsilon_{ij}$ mutually independent

- **Nature** ($A_{ij}$ and $D_{ij}$) versus nurture ($C_{ij}$ and $\epsilon_{ij}$)
  - **Heritability** is percentage of variance in trait that is due to genes
    \[ h^2 = \frac{\sigma_A^2 (+ \sigma_D^2)}{\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2} \]
**Liability model for binary traits**

- **Continuous ‘liability’ (propensity)**
  \[
y_{ij}^* = x_{ij}'\beta + A_{ij} + D_{ij} + C_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, 1)
\]

- **Binary trait**
  \[
y_{ij} = \begin{cases} 
1 & \text{if } y_{ij}^* > 0 \\
0 & \text{otherwise}
\end{cases}
\]

- **Probit model**
  \[
\Pr(y_{ij} = 1|x_{ij}, A_{ij}, D_{ij}, C_{ij}) = \Phi(x_{ij}'\beta + A_{ij} + D_{ij} + C_{ij})
\]
  \(\Phi(\cdot)\) is standard normal CDF (inverse probit link)

- **Heritability**
  \[
h^2 = \frac{\sigma_A^2(\sigma_D^2)}{\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + 1}
\]
Assumptions for models considered here

- Hardy-Weinberg equilibrium
- No epistasis (interactions between alleles at different loci)
- No gene-environment interactions
- Random (non-assortative) mating
- Correlations among error components
  - For $A_{ij}$ and $D_{ij}$ this follows from Mendelian genetics, under assumptions above, and from type of kinship
  - For $C_{ij}$ make additional assumptions
Model formulation

- Usually biometrical models for twin and family data expressed as a multi-group structural equation models (SEMs) and fitted in Mx, Mplus, or other SEM software.

- Can formulate models as mixed/multilevel models [Rabe-Hesketh, Gjessing & Skrondal, 2008] and fit them in Stata.
  - *xtmixed*: Continuous phenotypes and models that do not require equality constraints for variances at different levels.
  - *gllamm*: Continuous, binary (or ordinal) phenotypes.

- Models with the fewest random effects are easiest to estimate for binary (or ordinal) phenotypes.
Models for twin designs

- Monozygotic (MZ) or ‘identical’ twins share all genes by descent
- Dizygotic (DZ) or ‘fraternal’ twins share half their genes by descent
- Equal environment assumption: MZ and DZ twins have same degree of similarity in their environments, so that excess similarity between MZ twins can be attributed to the greater proportion of shared genes
Models for twin designs (cont’d)

Consider two twin pairs: (MZ1, MZ2), (DZ1, DZ2):

\[ \text{Cov}(A) = \sigma_A^2 \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 1/2 & 1 \end{bmatrix} \]

\[ \text{Cov}(D) = \sigma_D^2 \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1/4 \\ 0 & 0 & 1/4 & 1 \end{bmatrix} \]

\[ \text{Cov}(C') = \sigma_C^2 \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \]

\[ \text{Cov}(E) = \sigma_E^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

ACDE model not identified here; consider ACE and ADE (as well as AE, CE)
**Twin datasets**

- All data: \( M \) is dummy for MZ; \( \text{pair} \) is twin-pair \( j \); \( \text{member} \) is \( i \)
- Continuous adult heights `twin_bmi.dta` [Posthuma & Boomsma, 2005]
  - Variables: `height` (in cm) and `male`
  - 304 twin pairs (13% with height missing for one member)
    - 307 DZ members (40% male), 262 MZ members (43% male)
- Continuous neuroticism `twin_neur.dta` [Sham, 1998]
  - Variable: `neurot` (Eysenck personality questionnaire)
  - 794 female twin pairs (no missing data)
    - 272 DZ pairs, 522 MZ pairs
- Binary hay fever status `twin_hay.dta` [Hopper et al., 1990]
  - Variables: `h, male`, pair-level frequency weights `freq`
  - 3,807 twin pairs (no missing data)
    - 2,009 DZ pairs (18% male, 45% mix), 1,798 MZ pairs (31% male)
Parameterization 1 (P1) of ACE as mixed model

Three-level data

- Level 3: Twin-pair $j$
- Level 2: Member $i$ (same as level 1)

```
. use twin_bmi, clear
. list pair M member male height if s==1, sepby(pair) noobs

<table>
<thead>
<tr>
<th>pair</th>
<th>M</th>
<th>member</th>
<th>male</th>
<th>height</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>190</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>190.7</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>178</td>
</tr>
<tr>
<td>269</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>183</td>
</tr>
<tr>
<td>269</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>158.5</td>
</tr>
</tbody>
</table>
```

- Use twin-pair level (level-3) random effect $c_j^{(3)}$ with variance $\sigma_C^2$ for shared environment
- Use member level (level-1) residual $\epsilon_{ij}$ with variance $\sigma_E^2$ for unique environment
Parameterization 1 (P1) of ACE as mixed model (cont’d)

Problem: Additive genetic component completely shared (correlation 1) for MZ twins and partially shared (correlation 0.5) for DZ twins

Solution:

- Shared component $a_j^{(3)}$ with variance $\sigma_A^2$ contributes only half as much variance to DZ twins as to MZ twins

$$a_j^{(3)}[M_j + \sqrt{\frac{1}{2}M_j}]$$

- $M_j$ is dummy for MZ
- $\overline{M}_j = 1 - M_j$ is dummy for DZ

- Remaining variance for DZ twins comes from unshared component $a_{ij}^{(2)}$ with variance $\sigma_A^2$

$$a_{ij}^{(2)}\sqrt{\frac{1}{2}M_j}$$
Continuous adult height: P1 ACE

Cannot estimate in `xtmixed` because of equality constraint for variances at different levels

In `gllamm`:

```plaintext
generate var3 = M + sqrt(1/2) * (1-M)
generate var2 = sqrt(1/2) * (1-M)
eq var3: var3
eq var2: var2
generate one = 1
eq cons: one

cons def 1 [mem1_1]var2 = [pai2_1]var3
gllamm height male, i(member pair) nrf(1 2)
eqs(var2 var3 cons) nocor constr(1) adapt
```
**Continuous adult height: P1 ACE (cont’d)**

log likelihood = -1727.820312522015

( 1) \([\text{mem1}_1] \text{var2} - [\text{pai2}_1] \text{var3} = 0\)

|     | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----|--------|-----------|-------|------|----------------------|
| male | 12.99536 | .6166593 | 21.07 | 0.000 | 11.78673 14.20398   |
| _cons| 167.9549 | .438026  | 383.44| 0.000 | 167.0963 168.8134  |

Variance at level 1

2.392252 (.30445676)

Variances and covariances of random effects

***level 2 (member)

\(\text{var}(1): 40.342974 (5.1760754)\)

***level 3 (pair)

\(\text{var}(1): 40.342974 (5.1760754)\)
\(\text{cov}(2,1): \text{fixed at 0}\)
\(\text{var}(2): 1.8175006 (5.2567317)\)

\(. \text{disp} 40.342974/(40.342974+1.8175006+2.392252) \)
\(. 90551078\)
**Parameterization 2 (P2) of ACE as mixed model**

- Three-level model
  - Level 3: Twin-pair \( j \)
  - Level 2: Hybrid \( k \), \( k = \begin{cases} 
    \text{pair } j & \text{for MZ twins} \\
    \text{member } i & \text{for DZ twins}
  \end{cases} \)
  - Level 1: Member \( i \)
    - \( \epsilon_{ij} \) with variance \( \sigma_E^2 \) for unique environment as before
    - \( u_{kj}^{(2)} \) with variance \( \sigma_A^2/2 \) for half the additive genetic variance that is shared for MZ and unique for DZ
    - \( u_{j}^{(3)} \) with variance \( \sigma_A^2/2 + \sigma_C^2 \) for the other half of additive genetic variance that is shared for everyone and for common environment

**Note:** Only two random effects instead of three
Continuous adult height: P2 ACE

generate k = pair if M==1
replace k = member if M==0
xtmixed height male || pair: || k:, mle variance

|                | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------------|--------|-----------|-------|------|----------------------|
| height         |        |           |       |      |                      |
| male           | 12.99535 | 0.6150212 | 21.13 | 0.000 | 11.78993 14.20077    |
| _cons          | 167.9549 | 0.4379076 | 383.54| 0.000 | 167.0966 168.8131    |

Random-effects Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>pair: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>21.98932</td>
<td>3.482324</td>
<td>16.12173 29.99244</td>
</tr>
<tr>
<td>k: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>20.17123</td>
<td>2.588088</td>
<td>15.68621 25.9386</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>2.392253</td>
<td>0.3044573</td>
<td>1.864131 3.069997</td>
</tr>
</tbody>
</table>
Continuous adult height: P2 ACE: (cont’d)

- Already have \( \hat{\sigma}_E^2 \)
- Get \( \hat{\sigma}_A^2 \) and \( \hat{\sigma}_C^2 \) using `nlcom`

\[
\begin{align*}
\text{. nlcom (var}_A\text{: 2*exp(2*[lns2_1_1]_cons))} \\
> (\text{var}_C\text{: exp(2*[lns1_1_1]_cons)}-\exp(2*[lns2_1_1]_cons))
\end{align*}
\]

|               | Coef.       | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|---------------|-------------|-----------|------|-------|---------------------|
| var_A         | 40.34246    | 5.176177  | 7.79 | 0.000 | 30.19734 – 50.48758 |
| var_C         | 1.818089    | 5.256801  | 0.35 | 0.729 | -8.485051 – 12.12123|

\* Heritability:
. disp 40.34246/(40.34246+1.818089+2.392253)
\.90549771

- Use `_diparm` with option `ci(probit)` to get confidence interval for heritability; however, requires derivatives
- Would be nice to have `ci(probit)` option in `nlcom`!
Parameterization 2 for ACE, AE, ADE, CE

ACE: $\hat{\sigma}_A^2 = 2\hat{\text{Var}}(u_{kj}^{(2)})$ and $\hat{\sigma}_C^2 = \hat{\text{Var}}(u_k^{(3)}) - \hat{\text{Var}}(u_{kj}^{(2)})$

- Potential problem: $\hat{\sigma}_C^2$ can be negative
- Solution 1: AE: constrain $\sigma_C^2$ to zero by constraining $\text{Var}(u_j^{(3)}) = \text{Var}(u_{kj}^{(2)})$ (in gllamm only; see slide 22)
- Solution 2: ADE (see below)

ADE (same model as ACE): $\hat{\sigma}_A^2 = 3\hat{\text{Var}}(u_j^{(3)}) - \hat{\text{Var}}(u_{kj}^{(2)})$ and $\hat{\sigma}_D^2 = 2[\hat{\text{Var}}(u_{kj}^{(2)}) - \hat{\text{Var}}(u_j^{(3)})]$

CE: Set $\text{Var}(u_{kj}^{(2)}) = 0$, giving two-level model

Note: Conventional likelihood ratio tests to compare models are conservative [Dominicus et al., 2006]
Continuous neuroticism: P2 ADE

generate k = pair if M==1
replace k = member if M==0
xtmixed neurot || pair: || k:, mle variance

| neurot | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|------|------|----------------------|
| _cons  | 10.23203 | .1237788  | 82.66 | 0.000 | 9.989426 to 10.47463 |

Random-effects Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>pair: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>3.345268</td>
<td>1.034871</td>
<td>1.824351 to 6.134134</td>
</tr>
<tr>
<td>k: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>5.023933</td>
<td>1.187507</td>
<td>3.161151 to 7.984402</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>9.559881</td>
<td>.5823694</td>
<td>8.483966 to 10.77224</td>
</tr>
</tbody>
</table>
Continuous neuroticism: P2 ADE (cont’d)

Note that $\sigma^2_C = \text{Var}(u_k^{(3)}) - \text{Var}(u_{kj}^{(2)}) < 0$

For ADE model, get $\sigma^2_A$ and $\sigma^2_D$ using `nlcom`

```
. nlcom (var_A: 3*exp(2*[lns1_1_1]_cons) - exp(2*[lns2_1_1]_cons) )
>     (var_D: 2*(exp(2*[lns2_1_1]_cons) - exp(2*[lns1_1_1]_cons)))

var_A: 3*exp(2*[lns1_1_1]_cons) - exp(2*[lns2_1_1]_cons)
var_D: 2*(exp(2*[lns2_1_1]_cons) - exp(2*[lns1_1_1]_cons))
```

| neurot | Coef.   | Std. Err. | z      | P>|z|    | [95% Conf. Interval] |
|--------|---------|-----------|--------|--------|----------------------|
| var_A  | 5.01187 | 4.088337  | 1.23   | 0.220  | -3.001123 13.02486   |
| var_D  | 3.357331| 4.180764  | 0.80   | 0.422  | -4.836817 11.55148   |

* heritability
. disp (5.01187+3.357331)/(5.01187+3.357331+9.559881)
.46679473
### Binary hay fever status: P2 ADE

generate num3 = freq  
gllamm h male, i(k pair) link(probit) fam(binom)  
   adapt weight(num)

| h    | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|------|---------|-----------|-------|-------|----------------------|
| male | -0.1636205 | 0.0534943 | -3.06 | 0.002 | -0.2684675 to -0.0587736 |
| _cons| -0.6874611  | 0.040749  | -16.87 | 0.000 | -0.7673276 to -0.6075945 |

Variance and covariances of random effects

**Level 2 (k)**

- var(1): 0.89076163 (0.16434027)

**Level 3 (pair)**

- var(1): 0.65503535 (0.10341492)

---

Note: Estimation fast because only 40 rows of data and pair-level frequency weights
. nlcom (var_A: 3*[pair2]_cons^2 - [k1]_cons^2)
>   (var_D: 2*([k1]_cons^2 - [pair2]_cons^2))

var_A: 3*[pair2]_cons^2 - [k1]_cons^2
var_D: 2*([k1]_cons^2 - [pair2]_cons^2)

|       | Coef.    | Std. Err. |   z  | P>|z| | [95% Conf. Interval] |
|-------|----------|-----------|------|------|----------------------|
| var_A | 1.074344 | .3679161  | 2.92 | 0.003| .3532421 - 1.795447  |
| var_D | .4714526 | .4085908  | 1.15 | 0.249| -.3293708 - 1.272276 |

. *Heritability
. disp (1.074344+.4714526)/(1.074344+.4714526+1)
.60719564
Binary hay fever status: P2 AE (cont’d)

\[
\text{constr def 1} \quad [\text{pair2\_cons} = [k1\_cons]}
\]
\[
gllamm \ h \ \text{male, i(k pair) link(probit) fam(binom) adapt}
\]
\[
\text{weight(num) constr(1)}
\]

log likelihood = -4604.027077892745

( 1) - [k1\_cons + [pair2\_cons] = 0

| h    | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------|--------|-----------|-------|------|----------------------|
| male | -0.1608356 | 0.0523616 | -3.07 | 0.002 | -0.2634623 \to -0.0582088 |
| _cons| -0.6758232 | 0.0388389 | -17.40 | 0.000 | -0.751946 \to -0.5997004 |

Variances and covariances of random effects

***level 2 (k)

var(1): 0.73240456 (0.08174648)

***level 3 (pair)

var(1): 0.73240456 (0.08174648)

\[ \text{disp} \quad 0.73240456/(0.73240456+1) \]
\[ \quad 0.42276762 \]
ACE for nuclear family designs

Nuclear family with two children (mother, father, child1, child2)

\[
\text{Cov}(A) = \sigma_A^2 \begin{bmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1 \end{bmatrix}
\]

\[
\text{Cov}(C) = \sigma_C^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}
\]

\[
\text{Cov}(E) = \sigma_E^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\]
Parametrization as mixed model

- Four-level model
  - Level 4: Family \( k \)
  - Level 3: Hybrid: Sibling pair \( j \), individual parents \( i \)
  - Level 2: Member \( i \) (same as level 1)

\[
y_{ijk} = x'_{ik} \beta + a_{1k}^{(4)} [M_i + K_i/2] + a_{2k}^{(4)} [F_i + K_i/2] + a_{ijk}^{(2)} [K_i/\sqrt{2}] + c_{jk}^{(3)} + \epsilon_{ijk}
\]

- \( M_i \) is a dummy for mother, \( F_i \) for father, \( K_i \) for child
- \( \text{Var}(c_{jk}^{(3)}) = \sigma_C^2 \) and \( \text{Var}(\epsilon_{ijk}) = \sigma_E^2 \)
- First three terms represent additive genetic component with \( \text{Var}(a_{1k}^{(4)}) = \text{Var}(a_{2k}^{(4)}) = \text{Var}(a_{ijk}^{(2)}) = \sigma_A^2 \)
  - \( a_{1k}^{(4)} \) and \( a_{2k}^{(4)} \) induce the required additive genetic covariances between each parent and each child and among the children
  - \( a_{ijk}^{(2)} \) provides remaining variance \( \sigma_A^2/2 \) for children
Continuous birthweight: Nuclear family data

- 1000 Nuclear families from Norwegian birth registry [Magnus et al., 2001]
- One child per family (no level 3, \( j \)), model simplifies to two-level model

\[
y_{ijk} = x'_{ik} \beta + a_{1k}^{(4)} [M_i + K_i/2] + a_{2k}^{(4)} [F_i + K_i/2] + a_{ijk}^{(2)} K_i/\sqrt{2} + c_{jk}^{(3)} + \epsilon_{ijk}
\]

\[
y_{ik} = x'_{ik} \beta + a_{1k}^{(4)} [M_i + K_i/2] + a_{2k}^{(4)} [F_i + K_i/2] + a_{3k}^{(4)} K_i/\sqrt{2} + \epsilon_{ij}
\]

- Model with \( c_{jk}^{(3)} \) not identified
- \( a_{ijk}^{(2)} K_i/\sqrt{2} \equiv a_{3k}^{(4)} K_i/\sqrt{2} \) because \( K_i \) is non-zero for one member per family
- Level 4 becomes level 2

\[
y_{ik} = x'_{ik} \beta + a_{1k}^{(2)} [M_i + K_i/2] + a_{2k}^{(2)} [F_i + K_i/2] + a_{3k}^{(2)} K_i/\sqrt{2} + \epsilon_{ij}
\]
Continuous birthweight: Nuclear family data (cont’d)

- `fam_birthwt.dta` contains `M`, `F`, `K`, `family`, `bwt` and
  - `male`: dummy for being male
  - `first`: dummy for being the first child
  - `midage`: dummy for mother aged 20-35 at time of birth
  - `highage`: dummy for mother’s age above 35 at time of birth
  - `birthyr`: year of birth minus 1967

```
. list family M F K male birthyr bwt if family<3, sepby(family) noobs

<table>
<thead>
<tr>
<th>family</th>
<th>M</th>
<th>F</th>
<th>K</th>
<th>male</th>
<th>birthyr</th>
<th>bwt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3520</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>3940</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>3240</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3660</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3990</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>4330</td>
</tr>
</tbody>
</table>
```
Estimation using \textit{xtmixed}

Stata commands:

- \texttt{generate var1 = M + K/2}
- \texttt{generate var2 = F + K/2}
- \texttt{generate var3 = K/sqrt(2)}

\texttt{xtmixed bwt male first midage highage birthyr}
\begin{verbatim}
    || family: var1 var2 var3,
    nocons cov(ident) mle variance
\end{verbatim}

Note: Option \texttt{covariance(identity)} enforces variance equality constraint (and independence of error components) within a level
## Estimation using `xtmixed`

```
. xtmixed bwt male first midage highage birthyr || family: var1 var2 var3, 
>          nocons cov(ident) mle variance
```

|        | Coef.  | Std. Err. |    z  |    P>|z| | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|----------------------|
| bwt    |        |           |       |      |                      |
| male   | 158.4546 | 17.34853  | 9.13  | 0.000 | 124.4521 192.4571    |
| first  | -139.3974 | 18.7415  | -7.44 | 0.000 | -176.13 -102.6647   |
| midage | 57.0553  | 31.89569  | 1.79  | 0.074 | -5.459111 119.5697  |
| highage| 118.8564 | 54.67221  | 2.17  | 0.030 | 11.70082 226.0119   |
| birthyr| 3.627799 | .6882291  | 5.27  | 0.000 | 2.278894 4.976703   |
| _cons  | 3461.459 | 34.77956  | 99.53 | 0.000 | 3393.292 3529.625   |

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>family: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(var1 var2 var3)</td>
<td>99263.68</td>
<td>10157.96</td>
<td>81223.99 121310</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>133560.1</td>
<td>9069.929</td>
<td>116915.7 152574.2</td>
</tr>
</tbody>
</table>

LR test vs. linear regression: chibar2(01) = 97.80 Prob >= chibar2 = 0.0000
Concluding remarks

- Advantage of using multilevel models
  - More widely known and available in software than SEM
  - Can handle varying family sizes and missing data easily
  - Can extend to more levels, e.g., random neighborhood environment effects

- Other models considered in [Rabe-Hesketh, Skrondal & Gjessing, 2008]
  - Sibling and cousin data
  - Parameterization 1 for Twin ADE models

- Wishlist for Stata 12
  - Constraints for variance-covariance parameters in \texttt{xtmixed}, particularly equality constraints across levels
  - \texttt{nlcom} \texttt{with \texttt{ci(probit)}} option
References to own work


Other references


