LINEAR MIXED MODELS FOR LONGITUDINAL DATA

Estimating Individual Trajectories

Outline:

Linear mixed effects models for continuous outcomes

- Random Effect Estimation for U_i
- Variance estimation of $\hat{U}_i$
- Prediction and Shrinkage

- In mixed models, the random $\boldsymbol{U}_i$ for subject i can allow each subject to have his / her own intercept, slope (**w.r.t. time**) and even quadratic trend
  $\rightarrow$ The $\boldsymbol{U}_i$'s define **individual trajectories**

- Therefore, it may be of interest to **estimate** individual trajectories (i.e., the $\boldsymbol{U}_i$'s) for each subject, or for a subset of subjects

- **Excellent motivation** for using subject-specific mixed models over marginal models
  – you cannot get this with marginal models

- The mixed model is

$$Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i + Z_{ij}$$

  and in many applications we may consider Z_{ij} to be measurement error

- Therefore, it is of interest to estimate

$$\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i,$$

  attempting to eliminate the measurement error Z_{ij}

- For example, in the Riesby Depression data, we had that

$$b_{i0} = (\beta_0 + \beta_1 \texttt{endog}_i + U_{i1})$$

is the ith subject's **intercept** with respect to week, and

$$b_{i1} = (\beta_2 + \beta_3 \texttt{endog}_i + U_{i2})$$

is the ith subject's **slope** with respect to week

- The line $(b_{i0} + b_{i1}\texttt{week})$ for the average depression score for the ith subject over time is a function of $\boldsymbol{\beta}$ and $\boldsymbol{U}_i = (U_{i1}, U_{i2})'$

- Therefore, estimation of $\boldsymbol{U}_i$ is of interest

Estimation of the Random Effects $\boldsymbol{U}_i$: Frequentist Approach

- Estimation of $\boldsymbol{U}_i$ is not the same as estimation of $\boldsymbol{\beta}$ or of other parameters:
  - we have a lot of information on $\boldsymbol{\beta}$; only a little on each $\boldsymbol{U}_i$
  - $\boldsymbol{U}_i$ is modelled as a random variable (eg, $\boldsymbol{U}_i \sim N(0, G)$): Because there are many $\boldsymbol{U}_i$'s, it makes sense to consider them as random variables with a distribution
- Recall,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + D_i \boldsymbol{U}_i + \boldsymbol{Z}_i$$

  where

$$\boldsymbol{U}_i \sim N(0, G) \quad \text{and} \quad \boldsymbol{Z}_i \sim N(0, R_i)$$

  so $\boldsymbol{Y}_i | \boldsymbol{X}_i \sim N(\boldsymbol{X}_i \boldsymbol{\beta}, D_i G D_i' + R_i)$, likelihood is

$$\Pi_{i=1}^{m} f(\boldsymbol{Y}_i|\boldsymbol{\beta}, G, R_i)$$

and the log-likelihood does not have U_i

$$L(\boldsymbol{\beta}, G, R) = \sum_{i=1}^{m} \left\{ -\frac{1}{2} n_i \log(2\pi) - \frac{1}{2} \log |V_i| - \frac{1}{2} (\boldsymbol{Y}_i - X_i \boldsymbol{\beta})' V_i^{-1} (\boldsymbol{y} - X_i \boldsymbol{\beta}) \right\}$$

where $V_i = D_i G D_i' + R_i$.

- However, we can include U_i in log-likelihood (in form of conditional model) and obtain MLE for U_i:

$$\boldsymbol{Y}_i|(\boldsymbol{X}_i, U_i) \sim N(\boldsymbol{X}_i \boldsymbol{\beta} + D_i \boldsymbol{U}_i, R_i)$$

$$\boldsymbol{U}_i \sim N(0, G)$$

Then likelihood is

$$\Pi_{i=1}^{m} f(\boldsymbol{Y}_i|\boldsymbol{\beta}, U_i, R_i) f(U_i|G)$$

and the log-likelihood is

$$
\begin{aligned}
& L(\boldsymbol{\beta}, G, R) \\
= \quad & \sum_{i=1}^{m}\{-\frac{1}{2}\log|R_i| - \frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - D_i\boldsymbol{U}_i)' R_i^{-1}(\boldsymbol{Y} - \boldsymbol{X}_i\boldsymbol{\beta} - D_i\boldsymbol{U}_i) \\
& -\frac{1}{2}\log|G| - \frac{1}{2}U_i' G^{-1} U_i\} + \texttt{Constant}
\end{aligned}
$$

- Solve the score equation for U_i:

$$
\frac{\partial L(\boldsymbol{\beta}, G, R)}{\partial U_i} = D_i' R_i^{-1}(\boldsymbol{Y} - \boldsymbol{X}_i\boldsymbol{\beta} - D_i\boldsymbol{U}_i) - G^{-1}U_i = 0
$$

we have ML estimator for U_i:

$$
\widehat{U}_i = (G^{-1} + D_i' R_i^{-1} D_i)^{-1} D_i' R_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta})
$$

- In practice: $\boldsymbol{\beta}$, G, R_i are unknown
  $\rightarrow$ estimate $\boldsymbol{\beta}$, G, R_i via ML, ReML, etc.
  $\rightarrow$ estimate $\boldsymbol{U}_i$ as

$$\widehat{U_i} = \widehat{\mathrm{E}}(\boldsymbol{U}_i | \boldsymbol{Y}_i) = (\widehat{G}^{-1} + D_i' \widehat{R}_i^{-1} D_i)^{-1} D_i' \widehat{R}_i^{-1} (\boldsymbol{Y}_i - X_i \widehat{\boldsymbol{\beta}})$$

- This estimator is Best Linear Unbiased Predictor (BLUP) for U_i (Searle. et al. 1992) if G and R_i are known.

Estimation of the Random Effects $\boldsymbol{U}_i$: Bayesian Approach

- Bayes' Estimation: estimate $\boldsymbol{U}_i$ by $\mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i)$
  - $\boldsymbol{Y}_i$ is observed
  - so, why not estimate $\boldsymbol{U}_i$ as the **expected value** of $\boldsymbol{U}_i$ given what we observe, that is, given $\boldsymbol{Y}_i$
  - (**Prior**) the probability density of $\boldsymbol{U}_i$ is

$$f_u(\boldsymbol{U}_i)$$

  - the probability density of $\boldsymbol{Y}_i$ **given** $\boldsymbol{U}_i$ is

$$f_{y|u}(\boldsymbol{Y}_i|\boldsymbol{U}_i)$$

    called the **likelihood function** for $\boldsymbol{U}_i$ given data $\boldsymbol{Y}_i$

– the **marginal** distribution of $\boldsymbol{Y}_i$ averages out the $\boldsymbol{U}_i$

$$f_y(\boldsymbol{Y}_i) = \int f_y(\boldsymbol{Y}_i|\boldsymbol{U}_i) f_u(\boldsymbol{U}_i) \mathrm{d}\boldsymbol{U}_i$$

– **Posterior** distribution of $\boldsymbol{U}_i$ given $\boldsymbol{Y}_i$ via Bayes' theorem:

$$f_{u|y}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = \frac{f_{y|u}(\boldsymbol{Y}_i|\boldsymbol{U}_i) f_u(\boldsymbol{U}_i)}{f_y(\boldsymbol{Y}_i)}$$

– when we have the posterior, we can estimate $\boldsymbol{U}_i$ as

$$\mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i)$$

and this is called a **Bayes' estimate**

Key idea: "Best guess" at $\boldsymbol{U}_i$ given subject i's data, $\boldsymbol{Y}_i$

- So, lets do it for random effects
- Recall that

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + D_i\boldsymbol{U}_i + \boldsymbol{Z}_i$$

where

$$\boldsymbol{U}_i \sim N(0, G) \text{ and } \quad \boldsymbol{Z}_i \sim N(0, R_i)$$

- we have

$$\mathrm{E}(\boldsymbol{Y}_i) = X_i\boldsymbol{\beta}$$

and that

$$\mathrm{var}(\boldsymbol{Y}_i) = V_i = D_iGD_i' + R_i \qquad \text{(often } R_i = \tau^2 I_{n_i}\text{)}$$

- Also note that

$$\mathrm{cov}(\boldsymbol{Y}_i, \boldsymbol{U}_i) = \mathrm{cov}(D_i\boldsymbol{U}_i, \boldsymbol{U}_i) = D_i\mathrm{var}(\boldsymbol{U}_i) = D_iG$$

- Therefore,

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{U}_i \end{pmatrix} \sim N\left[\begin{pmatrix} X_i\boldsymbol{\beta} \\ 0 \end{pmatrix}, \begin{pmatrix} V_i & D_iG \\ GD_i' & G \end{pmatrix}\right]$$

- We can obtain posterior of $\boldsymbol{U}_i$ given $\boldsymbol{Y}_i$:

$$\boldsymbol{U}_i|\boldsymbol{Y}_i \sim N[GD_i'V_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}), G - GD_i'V_i^{-1}D_iG]$$

- Above derivation used following result (as part of your basic "statistician's toolbox"):
  If

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

  then

$$\boldsymbol{x}_1|\boldsymbol{x}_2 \sim N[\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]$$

- Thus, we can obtain **Bayes' estimate of $\boldsymbol{U}_i$**:

$$\begin{aligned} \mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i) &= G D_i' V_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}) \\ &= G D_i'(D_i G D_i' + R_i)^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}) \end{aligned}$$

- With some math work we can show

$$\mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = (G^{-1} + D_i' R_i^{-1} D_i)^{-1} D_i' R_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta})$$

  - Same form as derived previously using frequentist approach
  - To see this, multiply $(G^{-1} + D_i' R_i^{-1} D_i)$ and $(D_i G D_i' + R_i)$ to following equation, so you can see it hold:

$$G D_i'(D_i G D_i' + R_i)^{-1} = (G^{-1} + D_i' R_i^{-1} D_i)^{-1} D_i' R_i^{-1}$$

- In practice, estimate $\boldsymbol{U}_i$ as

$$\widehat{U_i} = \widehat{\mathrm{E}}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = (\widehat{G}^{-1} + D_i'\widehat{R}_i^{-1}D_i)^{-1}D_i'\widehat{R}_i^{-1}(\boldsymbol{Y}_i - X_i\widehat{\boldsymbol{\beta}})$$

  – $\widehat{\boldsymbol{U}}_i$ is called an **empirical Bayes' estimate** because $\boldsymbol{\beta}$, G, R_i are based on data rather than known

  – it is the **estimated** expected value of $\boldsymbol{U}_i$ given $\boldsymbol{Y}_i$
  (Our "best guess" at $\boldsymbol{U}_i$ given estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{G}$, $\widehat{R}$ and subject i's data, $\boldsymbol{Y}_i$)

Note:

- Above results involve unknown parameters $\boldsymbol{\beta}$, G, R_i
- Empirical Bayes estimation: substitute point estimates for these parameters in posterior distributions.
- A full Bayesian inference: should assume prior distributions for them, and integrate these parameters out to obtain distribution of $\boldsymbol{U}_i|\boldsymbol{Y}_i$

Variance Estimation of $\widehat{\boldsymbol{U}}_i$

- Based on posterior distribution, we have

$$\mathrm{var}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = G - GD_i'(D_iGD_i' + R_i)^{-1}D_iG$$

- With some math work you can prove

$$\mathrm{var}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = (G^{-1} + D_i'R_i^{-1}D_i)^{-1}$$

due to the fact that

$$(A - BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B - D^{-1})^{-1}B'A^{-1}$$

(plug in G^{-1} for A, $-R_i^{-1}$ for D, and D_i' for B),

- (Naive) Variance of $\widehat{U_i}$ can be estimated by:

$$\widehat{\mathrm{var}}_{\beta,G,R_i}(\boldsymbol{U}_i|\boldsymbol{Y}_i) = (\hat{G}^{-1} + D_i'\hat{R}_i^{-1}D_i)^{-1}$$

  $\widehat{\mathrm{var}}_{\beta,G,R_i}$ means that it is conditional on fixed values of β, G, R_i

- However, this variance does not account for the variability in the estimation of $\boldsymbol{\beta}$, G and R_i in the calculation of $\mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i)$

- A full Bayesian approach: take into account the estimation of $\boldsymbol{\beta}$, G and R_i
  – Assign priors to the parameters $\boldsymbol{\beta}$, G and R_i (ie, τ^2 if $R_i = \tau^2 I_{n_i}$)
  – calculate $\mathrm{var}(\boldsymbol{U}_i|\boldsymbol{Y}_i)$ integrating out $\boldsymbol{\beta}$, G and R_i

- We use an 'intermediate' Bayesian calculation:
  - fix G and R_i at their estimated value
  - assign $\boldsymbol{\beta}$ a flat prior (recall: this uninformative prior will lead to same inference of $\boldsymbol{\beta}$ with the one obtained by OLS/WLS)

- (Integrating out $\boldsymbol{\beta}$) we have

$$
\begin{aligned}
& \operatorname{var}_{G,R_i}(\boldsymbol{U}_i|\boldsymbol{Y}_i) \\
= \; & \mathrm{E}_{G,R_i}\left[\operatorname{var}_{\beta,G,R_i}(\boldsymbol{U}_i|\boldsymbol{Y}_i)\right] + \operatorname{var}_{G,R_i}\left[\mathrm{E}_{\beta,G,R_i}(\boldsymbol{U}_i|\boldsymbol{Y}_i)\right] \\
= \; & \mathrm{E}_{G,R_i}\left[(G^{-1} + D_i' R_i^{-1} D_i)^{-1}\right] + \operatorname{var}_{G,R_i}\left[G D_i' V_i^{-1}(\boldsymbol{Y}_i - X_i \boldsymbol{\beta})\right] \\
= \; & (G^{-1} + D_i' R_i^{-1} D_i)^{-1} \\
& + G D_i' V_i^{-1} X_i \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1} X_i' V_i^{-1} D_i G
\end{aligned}
$$

where we used the result of

$$
\operatorname{var}(\boldsymbol{\beta}) = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1}
$$

obtained by WLS (Note 5)

- This formula is preferred to estimate variance of $\widehat{U_i}$

Special Case: Empirical Bayes' Estimation of a Random Intercept

- For a simple case of just a **random intercept** U_i:

$$G = \nu^2 = \mathrm{var}(\boldsymbol{U}_i) \quad \text{(a scalar)}$$

and, for only a random intercept

$$D_i = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and} \quad R_i = \tau^2 I_{n_i} = \begin{pmatrix} \tau^2 & 0 & \cdots & 0 \\ 0 & \tau^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau^2 \end{pmatrix} = \mathrm{var}(\boldsymbol{Z}_i)$$

- For the ith subject: If $n_i =$ the number of observations, $\bar{Y}_i$ is the mean response (average within ith subject), and $\bar{\boldsymbol{x}}_i$ is the mean predictor vector, then

$$D_i' R_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}) = \frac{n_i}{\tau^2}(\bar{Y}_i - \bar{\boldsymbol{x}}_i\boldsymbol{\beta})$$

and

$$G^{-1} = \frac{1}{\nu^2}$$

and

$$D_i' R_i^{-1} D_i = \frac{n_i}{\tau^2}$$

from which we obtain

$$\mathrm{E}(U_i|\boldsymbol{Y}_i) = \frac{1}{\left(\frac{1}{\nu^2} + \frac{n_i}{\tau^2}\right)}\left(\frac{n_i}{\tau^2}\right)(\bar{Y}_i - \bar{\boldsymbol{x}}_i\boldsymbol{\beta}) = \widehat{\boldsymbol{U}}_i,$$

which becomes an **empirical Bayes'** estimate when we replace $\boldsymbol{\beta}$, ν^2, τ^2 with estimates

- Define the **weight**:

$$w_i = \frac{\left(\frac{n_i}{\tau^2}\right)}{\left(\frac{1}{\nu^2} + \frac{n_i}{\tau^2}\right)}, \quad 0 < w_i < 1$$

- Now, suppose we wanted to estimate the intercept for the ith subject:

$$b_{0i} = \beta_0 + U_i$$

- We would estimate $\widehat{\boldsymbol{\beta}}$, $\widehat{\nu}^2$, $\widehat{\tau}^2$, and then compute the empirical Bayes estimate of b_{0i}

$$\hat{b}_{0i} = \widehat{w}_i(\bar{Y}_i - \bar{\boldsymbol{x}}_i\widehat{\boldsymbol{\beta}}) + \widehat{\beta}_0 = \widehat{w}_i(\bar{Y}_i - \bar{\boldsymbol{x}}_i\widehat{\boldsymbol{\beta}} + \widehat{\beta}_0) + (1 - \widehat{w}_i)\widehat{\beta}_0$$

- Now, think about w_i:
  - the smaller w_i is, the closer $\hat{b}_{0i}$ is to $\widehat{\beta}_0$

- w_i "shrinks" the estimate of the individual intercept b_{0i} towards the population average intercept β_0
- If n_i is big (many observations on subject i)

  or

  if τ^2 is small (low within-subject variance)

  then w_i is large, indicating that $\boldsymbol{Y}_i$ is **very informative** for $b_{0i} = \beta_0 + U_i$
- If ν^2 is small, then there is very little between-subject variance of intercepts, indicating that the population mean intercept β_0 is a pretty good estimate of the subject-specific intercept $b_{0i} = \beta_0 + U_i$
- In this sense, the total information about $\beta_0 + U_i$ is

$$\left(\frac{1}{\nu^2} + \frac{n_i}{\tau^2}\right)$$

  and w_i creates the **optimal weighting** among the two source of information

- Empirical Bayes' estimator **shrinks** the estimate of b_{0i} toward the population average β_0 via the weights $\widehat{w}_i$

  Sometimes they are called **shrinkage estimators**

- Empirical Bayes' estimators are optimally weighted to achieve the lowest mean-square error of $\hat{b}_{0i}$ across all of the subjects

- They do this by **borrowing strength** (information) from other subjects via $\widehat{\beta}_0$ in estimating b_{0i}

- if we did not borrow this information, we would estimate b_{0i} using only the data from subject i

$$(\bar{y}_i - \bar{\boldsymbol{x}}_i \boldsymbol{\beta})$$

Shrinkage in general mixed-effects model

- ith subject's predicted response is

$$\begin{aligned}
\widehat{\boldsymbol{Y}}_i &= X_i\widehat{\boldsymbol{\beta}} + D_i\widehat{U_i} \\
&= X_i\widehat{\boldsymbol{\beta}} + D_i\widehat{G}D_i'\widehat{V}_i^{-1}(\boldsymbol{Y}_i - X_i\widehat{\boldsymbol{\beta}}) \\
&= (I_{n_i} - D_i\widehat{G}D_i'\widehat{V}_i^{-1})X_i\hat{\boldsymbol{\beta}} + D_i\widehat{G}D_i'\widehat{V}_i^{-1}\boldsymbol{Y}_i \\
&= (\widehat{R}_i\widehat{V}_i^{-1})X_i\widehat{\boldsymbol{\beta}} + (I_{n_i} - \widehat{R}_i\widehat{V}_i^{-1})\boldsymbol{Y}_i
\end{aligned}$$

  This is due to fact that

$$I_{n_i} = \widehat{V}_i\widehat{V}_i^{-1} = (D_i\widehat{G}D_i' + \widehat{R}_i)\widehat{V}_i^{-1} = D_i\widehat{G}D_i'\widehat{V}_i^{-1} + \widehat{R}_i\widehat{V}_i^{-1}$$

- Empirical Bayes estimator for the ith subject's predicted response is a weighted average of the individual response $\boldsymbol{Y}_i$ and the population-averaged mean response $X_i\widehat{\boldsymbol{\beta}}$.

- It "shrinks" the subject's response profile toward the population mean profile.
- Amount of "shrinkage" depends on the relative magnitude of R_i (ie, within subject variability) and $V_i = D_i G D_i' + R_i$ (total variation)
- If within subject variability (R_i) is large relative to between-subject variability $(D_i G D_i')$, more weight is assigned to the overall average profile
- more weight is given to the observed data Y_i if the opposite is true.

Riesby Depression Example: Empirical Bayes Estimation of Individual Trajectories

- Empirical Bayes estimation of individual random effects $\boldsymbol{U}_i$ are available in many software packages

- **Predicted trajectories** is

$$\widehat{Y}_{ij} = \boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{d}'_{ij}\widehat{\boldsymbol{U}}_{ij}$$

- SAS code to obtain predicted trajectories:

```
data riesby;
set riesby;
endwk=endog*week;
wkcen = week-2.5;
endwkcen = endog*wkcen;
wksqr = wkcen*wkcen;
run;
```

```
ods output solutionR=riesby.randeff;
proc mixed data=riesby covtest;
class id;
model hamd=endog week endwk/ s outp=riesby.predict outpm=riesby.popmean;
random intercept week / subject=id type=un g solution;
run;

proc print data=riesby.predict (obs=10);
title1 'predicted individual data';
run;
```

- "solution" option in RANDOM statement: estimate the random effects U_i
- "outp=" option: specifies a SAS data set which contains predicted mean value $\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{d}'_{ij}\widehat{\boldsymbol{U}}_{ij}$
- "outpm=" option: specifies a SAS data set that contains population mean estimates $\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}}$

- “ods output solutionR=” option: specifies a SAS data set that contains the random effect estimates $\widehat{\boldsymbol{U}}_i$

```
                          predicted individual data

Obs           id           hamd            week          endog    endwk

  1          101             26               0              0        0
  2          101             22               1              0        0
  3          101             18               2              0        0
  4          101              7               3              0        0
  5          101              4               4              0        0
  7          103             33               0              0        0
  8          103             24               1              0        0
  9          103             15               2              0        0
 10          103             24               3              0        0

                     StdErr
Obs       Pred         Pred      DF    Alpha      Lower      Upper      Resid

  1    24.3368      2.02843     243     0.05    20.3413    28.3324    1.66319
  2    19.9049      1.56921     243     0.05    16.8139    22.9959    2.09510
  3    15.4730      1.32719     243     0.05    12.8587    18.0873    2.52700
```

```
  4    11.0411    1.41831    243      0.05      8.2473    13.8348    -4.04109
  5     6.6092    1.79245    243      0.05      3.0785    10.1399    -2.60919
  6     2.1773    2.31631    243      0.05     -2.3853     6.7399     0.82272
  7    26.8659    2.02843    243      0.05     22.8704    30.8615     6.13409
  8    24.0879    1.56921    243      0.05     20.9969    27.1789    -0.08788
 10    18.5318    1.41831    243      0.05     15.7381    21.3256     5.46819
```

```
proc print data=riesby.popmean (obs=10);
title1 'population mean';
run;
```

```
                              population mean

Obs              id            hamd             week             endog    endwk

  1             101              26                0                 0        0
  2             101              22                1                 0        0
  3             101              18                2                 0        0
  4             101               7                3                 0        0
  5             101               4                4                 0        0
  6             101               3                5                 0        0
  7             103              33                0                 0        0
  8             103              24                1                 0        0
```

9	103	15	2	0	0
10	103	24	3	0	0

Obs	Pred	StdErr Pred	DF	Alpha	Lower	Upper	Resid
1	22.4760	0.80742	243	0.05	20.8856	24.0664	3.5240
2	20.1103	0.72196	243	0.05	18.6882	21.5324	1.8897
3	17.7446	0.76909	243	0.05	16.2297	19.2595	0.2554
4	15.3789	0.92885	243	0.05	13.5493	17.2085	-8.3789
5	13.0132	1.15542	243	0.05	10.7373	15.2891	-9.0132
6	10.6475	1.41712	243	0.05	7.8561	13.4389	-7.6475
7	22.4760	0.80742	243	0.05	20.8856	24.0664	10.5240
8	20.1103	0.72196	243	0.05	18.6882	21.5324	3.8897
9	17.7446	0.76909	243	0.05	16.2297	19.2595	-2.7446
10	15.3789	0.92885	243	0.05	13.5493	17.2085	8.6211

```
proc print data=riesby.randeff (obs=10);
title1 'random coefficients';
run;
```

random coefficients

Obs	Effect	id	Estimate	StdErr Pred	DF	tValue	Probt

1	Intercept	101	1.8608	2.0781	243	0.90	0.3714
2	week	101	-2.0662	0.7259	243	-2.85	0.0048
3	Intercept	103	4.3899	2.0781	243	2.11	0.0357
4	week	103	-0.4123	0.7259	243	-0.57	0.5706
5	Intercept	104	2.0398	2.0647	243	0.99	0.3242
6	week	104	-1.5457	0.7179	243	-2.15	0.0323
7	Intercept	105	-2.2202	2.0781	243	-1.07	0.2864
8	week	105	0.2774	0.7259	243	0.38	0.7027
9	Intercept	106	-0.3326	2.1092	243	-0.16	0.8748
10	week	106	0.9878	0.8646	243	1.14	0.2544

- Note: In Stata, predicted trajectories are generated after the model fit with the `predict` command:

```
. xtmixed hamd endog week endwk || id: week , cov(uns) , var
. predict pred , fitted
```

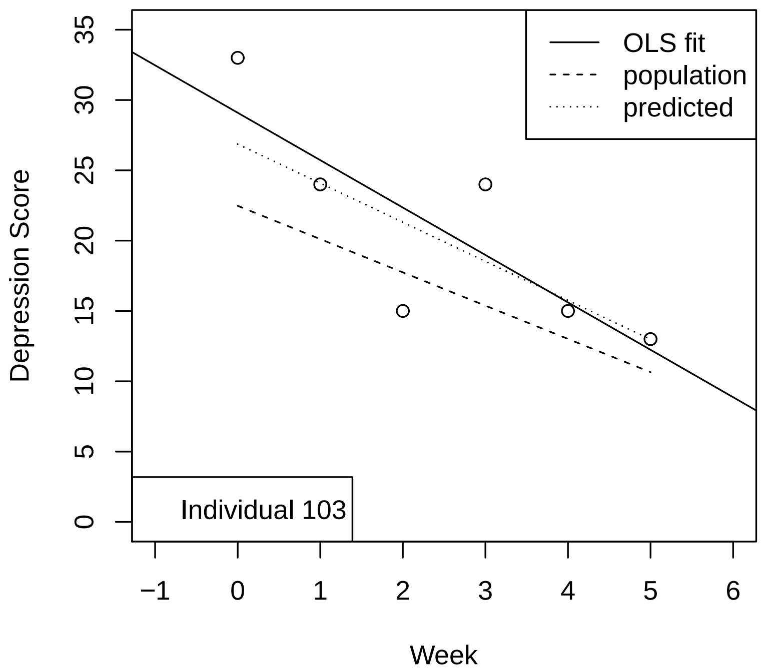
- To see how the predicted trajectories fit the data, we plot the trajectories in R:

```
#Input SAS data sets
library(foreign)
sashome <- "E:/Program Files/SASHome/SASFoundation/9.4"

datapath="C:/Users/Shuai Chen/OneDrive - University of California, Davis/teach/lon
data1<-read.ssd(datapath,"randeff",sascmd=file.path(sashome,"sas.exe"))
data2<-read.ssd(datapath,"popmean",sascmd=file.path(sashome,"sas.exe"))
data3<-read.ssd(datapath,"predict",sascmd=file.path(sashome,"sas.exe"))

#Plots for individual, eg, ID=103
id<-103
plot(data2$WEEK[data2$ID==id],data2$HAMD[data2$ID==id], xlim=c(-1,6),ylim=c(0,35)
legend("topright",c("OLS fit","population","predicted"),lty=1:3)
legend("bottomleft", paste("Individual",id))
abline(lsfit(data2$WEEK[data2$ID==id],data2$HAMD[data2$ID==id]))
lines(data2$WEEK[data2$ID==id],data2$PRED[data2$ID==id],lty=2)
lines(data3$WEEK[data3$ID==id],data3$PRED[data3$ID==id],lty=3)
```

- Fitted trajectories: random intercept and random slope:

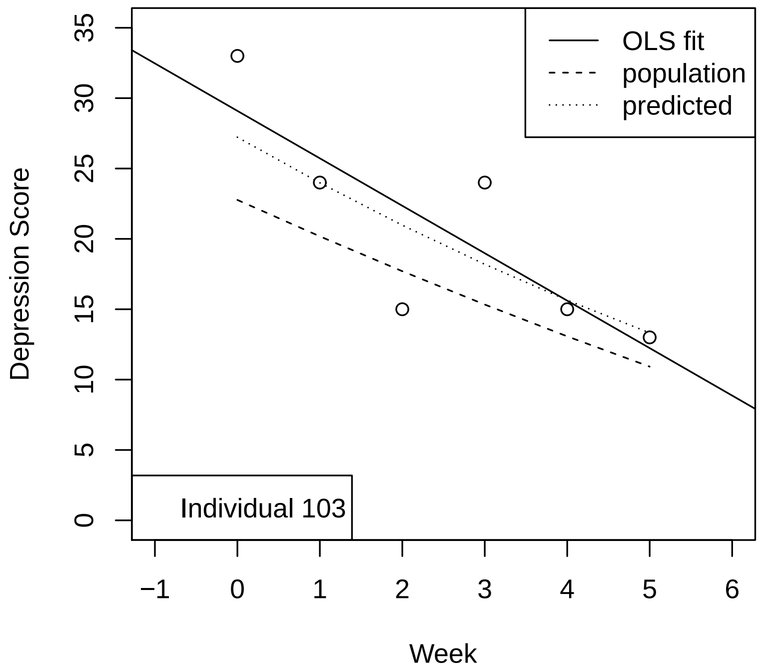


- OLS: use only ith subject's observations

- Then, for comparison, we did the same thing for the model with a quadratic random effect, specifically

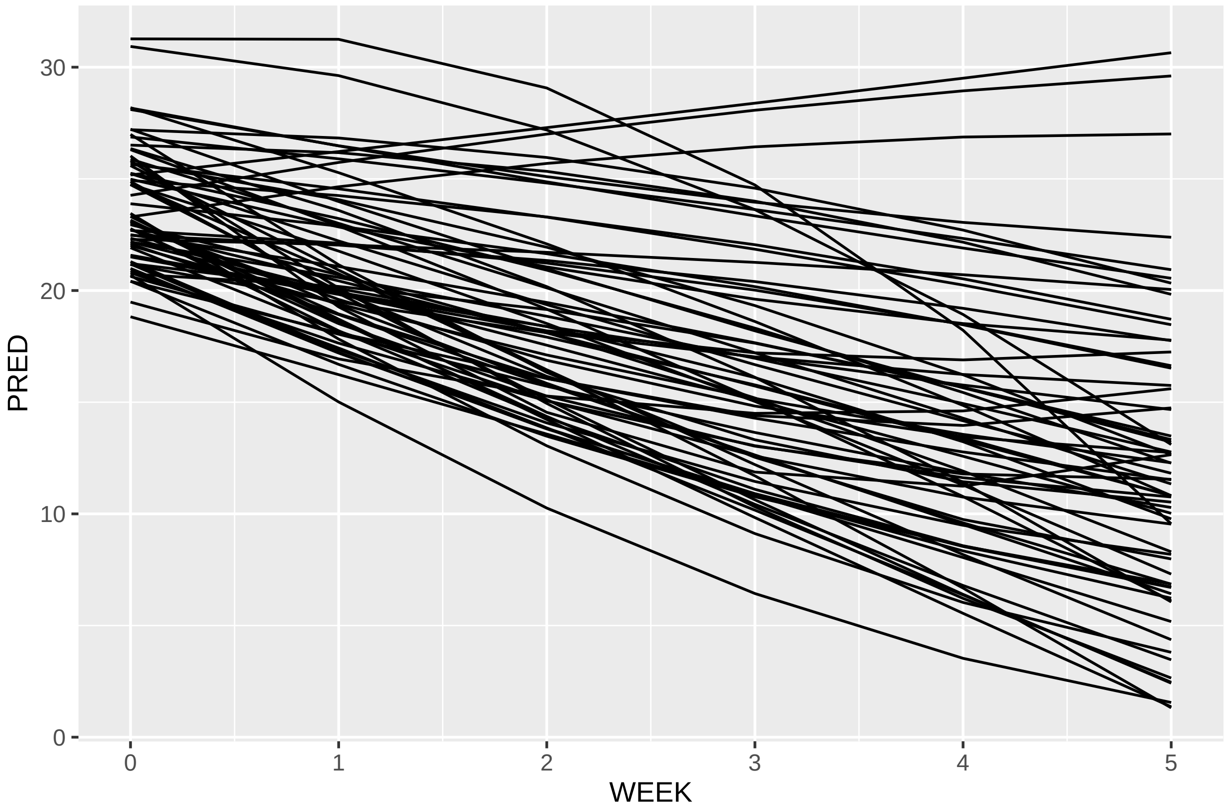
```
ods output solutionR=riesby.randeff2;
proc mixed data=riesby covtest;
class id;
model hamd=endog wkcen endwkcen wksqr/ s outp=riesby.predict2
     outpm=riesby.popmean2;
random intercept wkcen wksqr/ subject=id type=un g solution;
run;
```

- Fitted trajectories: random intercept, slope and quadratic term:



- Notes:
  - the trajectories for subjects who are overall high or overall low are "shrunk" toward the middle
  - individual points are also shrunk towards the middle
  - the fitted trajectories include all points, even when data are missing
  - when the quadratic term is added, it captures the non-linear trends better for those subjects who have them
  - to get an overall idea of how much non-linearity the data exhibit across subjects, plot the trajectories for all subjects (with quadratic random effects):

    ```
    ggplot(data = data3, aes(x = WEEK, y = PRED, group = ID))+ geom_line()
    ```



– a few subjects have much more pronounced non-linearities in their trajectories

– significance of the quadratic random effect may be only due to these few subjects, a possibility worth investigating