Bilinear random effects models for matrix data

Peter Hoff
Duke STA 610

Matrix-valued data

Data matrix: $\mathbf{Y} \in \mathbb{R}^{m \times n}$.

- $\mathbf{Y} \in \mathbb{R}^{m \times n}$.
- $y_{i,j}$ = measurement specific to ith row unit, jth column unit.

Examples:

- Classical multivariate data: $y_{i,j}$ = jth variable for person i.
- Panel data: $y_{i,j}$ = outcome for unit i at time j (mv time series);
- Rating data: $y_{i,j}$ = rating of object i by person j.
- Dyadic data: $y_{i,j}$ = interaction between units i and j.

Matrix-valued data

Data matrix: $\mathbf{Y} \in \mathbb{R}^{m \times n}$.

- $\mathbf{Y} \in \mathbb{R}^{m \times n}$.
- $y_{i,j}$ = measurement specific to ith row unit, jth column unit.

Examples:

- Classical multivariate data: $y_{i,j}$ = jth variable for person i.
- Panel data: $y_{i,j}$ = outcome for unit i at time j (mv time series);
- Rating data: $y_{i,j}$ = rating of object i by person j.
- Dyadic data: $y_{i,j}$ = interaction between units i and j.

Indices as grouping factors

In some situations, the row and column indices may be though of as grouping factors.

- If the levels of an index set represent very different things (height, weight, income, education), then probably don't take this perspective.
- If they represent members of a common set of objects, then this perspective can be useful.

- a collection of survey participants, a collection of time points;
- a collecion of movie raters, a collection of movies;
- a collection of people in a social network.

Indices as grouping factors

In some situations, the row and column indices may be though of as grouping factors.

- If the levels of an index set represent very different things (height, weight, income, education), then probably don't take this perspective.
- If they represent members of a common set of objects, then this perspective can be useful.

- a collection of survey participants, a collection of time points;
- a collecion of movie raters, a collection of movies;
- a collection of people in a social network.

Additive random effects models

Basic additive random effects model:

$$y_{i,j} = \beta^\top x_{i,j} + a_i + b_j + \epsilon_{i,j}$$
$$a_1, \ldots, a_m \sim \text{ i.i.d } N(0, \tau_a^2)$$
$$b_1, \ldots, b_n \sim \text{ i.i.d } N(0, \tau_b^2).$$

In a homework, we showed that this implies

$$\text{Cov}[y_{i,j}, y_{i',j'}] = \tau_a^2 \text{ if } i = i'$$
$$\text{Cov}[y_{i,j}, y_{i',j'}] = \tau_b^2 \text{ if } j = j'$$
$$\text{Cov}[y_{i,j}, y_{i',j'}] = 0 \text{ if } i \neq i', j \neq j' \ .$$

For example, for panel data this implies exchangeability among all observations within a common time point (they are all equally correlated).

Additive random effects models

Basic additive random effects model:

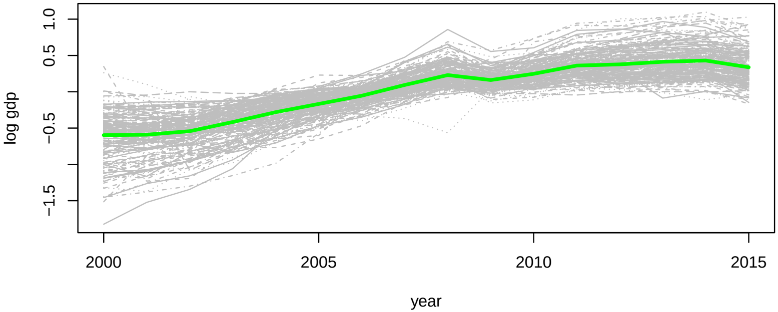
$$y_{i,j} = \beta^\top x_{i,j} + a_i + b_j + \epsilon_{i,j}$$
$$a_1, \ldots, a_m \sim \text{ i.i.d } N(0, \tau_a^2)$$
$$b_1, \ldots, b_n \sim \text{ i.i.d } N(0, \tau_b^2).$$

In a homework, we showed that this implies

$$\text{Cov}[y_{i,j}, y_{i',j'}] = \tau_a^2 \text{ if } i = i'$$
$$\text{Cov}[y_{i,j}, y_{i',j'}] = \tau_b^2 \text{ if } j = j'$$
$$\text{Cov}[y_{i,j}, y_{i',j'}] = 0 \text{ if } i \neq i', j \neq j' \ .$$

For example, for panel data this implies exchangeability among all observations within a common time point (they are all equally correlated).

Example: Per-capita GDP



Simple additive fit

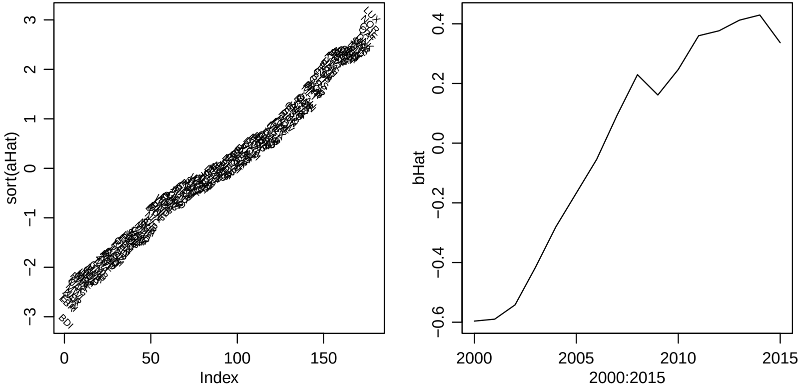
```
xrow<-rep( rownames(Y) , times=ncol(Y))
xcol<-rep( colnames(Y) , times=rep(nrow(Y),ncol(Y)) )
y<-c(Y)

cbind( y,xrow,xcol)[1:10,]

##       y                  xrow  xcol
##  [1,] "7.06969467420624" "ALB" "2000"
##  [2,] "7.47136990344077" "DZA" "2000"
##  [3,] "6.40732921228798" "AGO" "2000"
##  [4,] "9.21977166366161" "ATG" "2000"
##  [5,] "8.94497719738542" "ARG" "2000"
##  [6,] "6.43201494839607" "ARM" "2000"
##  [7,] "9.98345866321392" "AUS" "2000"
##  [8,] "10.1071329419972" "AUT" "2000"
##  [9,] "6.48478397663778" "AZE" "2000"
## [10,] "9.96369634168077" "BHS" "2000"
```

Simple additive fit

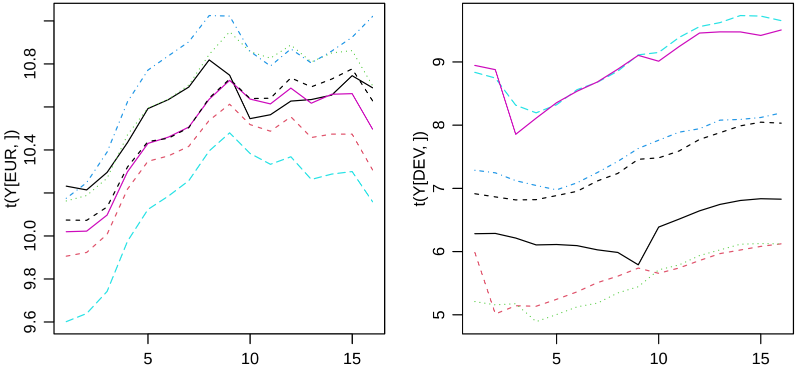
```
library(lme4)
fit<-lmer( y ~  1 + (1|xrow) + (1|xcol) )
summary(fit)

## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 + (1 | xrow) + (1 | xcol)
##
## REML criterion at convergence: 660.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6522 -0.5398  0.0014  0.5775  4.3365
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  xrow     (Intercept) 2.32474  1.5247
##  xcol     (Intercept) 0.14474  0.3804
##  Residual             0.04703  0.2169
## Number of obs: 2848, groups:  xrow, 178; xcol, 16
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.2542     0.1487   55.49
```

Estimated effects

```
aHat<-ranef(fit)[[1]][[1]] ; names(aHat)<-rownames(ranef(fit)[[1]] )
bHat<-ranef(fit)[[2]][[1]] ; names(bHat)<-rownames(ranef(fit)[[2]] )
```



Lack of additivity



Multiplicative models

Consider an "interaction:"

$$y_{i,j} = \mu + a_{i,1} + b_{j,1} + a_{i,2} b_{j,2} + \epsilon_{i,j}$$

- For countries with $a_{i,2} \approx 0$, the time trend is $\{b_{1,1}, \ldots, b_{1,T}\}$.
- For countries with $a_{i,2} \approx 1$, the time trend is $\{b_{1,1} + b_{2,1}, \ldots, b_{1,T} + b_{2,T}\}$.

Such a model allows for time trajectories that vary across units.

To accommodate more trajectories, we increase the number of interactions:

$$y_{i,j} = \mu + a_i + b_j + \mathbf{u}_i^\top \mathbf{v}_j + \epsilon_{i,j}$$
$$\mathbf{u}_i^\top \mathbf{v}_j = u_{i,1} v_{j,1} + \cdots + u_{i,r} v_{j,r}.$$

Multiplicative models

Consider an "interaction:"

$$y_{i,j} = \mu + a_{i,1} + b_{j,1} + a_{i,2}b_{j,2} + \epsilon_{i,j}$$

- For countries with $a_{i,2} \approx 0$, the time trend is $\{b_{1,1}, \ldots, b_{1,T}\}$.
- For countries with $a_{i,2} \approx 1$, the time trend is $\{b_{1,1} + b_{2,1}, \ldots, b_{1,T} + b_{2,T}\}$.

Such a model allows for time trajectories that vary across units.

To accommodate more trajectories, we increase the number of interactions:

$$y_{i,j} = \mu + a_i + b_j + \mathbf{u}_i^\top \mathbf{v}_j + \epsilon_{i,j}$$
$$\mathbf{u}_i^\top \mathbf{v}_j = u_{i,1}v_{j,1} + \cdots + u_{i,r}v_{j,r}.$$

AMMI models

Such models are called “AMMI models”:

- additive main effects;
- multiplicative interactions.

The basic structure is used widely in modern multivariate analysis:

- panel data;
- consumer ratings of products;
- recommender systems;
- network analysis;
- genomics and bioinformatics.

The models are nonlinear, but are the next easiest thing.

Bilinear estimation via linear methods

Consider iterative estimation:

Update column effects: Given row effects $(a_i, \mathbf{u}_i)$,

$$\begin{aligned} y_{i,j} &= \mu + (a_i, \mathbf{u}_i)^\top (b_j, \mathbf{v}_j) + \epsilon_{i,j} \\ &\equiv \mu + \mathbf{x}_i^\top \beta_j + \epsilon_{i,j} \end{aligned}$$

which is *linear* in the column effects $(b_j, \mathbf{v}_j)$.

Similarly,
Update row effects: Given column effects $(b_j, \mathbf{v}_j)$,

$$\begin{aligned} y_{i,j} &= \mu + (a_i, \mathbf{u}_i)^\top (b_j, \mathbf{v}_j) + \epsilon_{i,j} \\ &\equiv \mu + \beta_i^\top \mathbf{x}_j + \epsilon_{i,j} \end{aligned}$$

which is *linear* in the row effects (a_i, u_j).

Bilinear estimation via linear methods

Consider iterative estimation:

Update column effects: Given row effects $(a_i, \mathbf{u}_i)$,

$$\begin{aligned} y_{i,j} &= \mu + (a_i, \mathbf{u}_i)^\top (b_j, \mathbf{v}_j) + \epsilon_{i,j} \\ &\equiv \mu + \mathbf{x}_i^\top \beta_j + \epsilon_{i,j} \end{aligned}$$

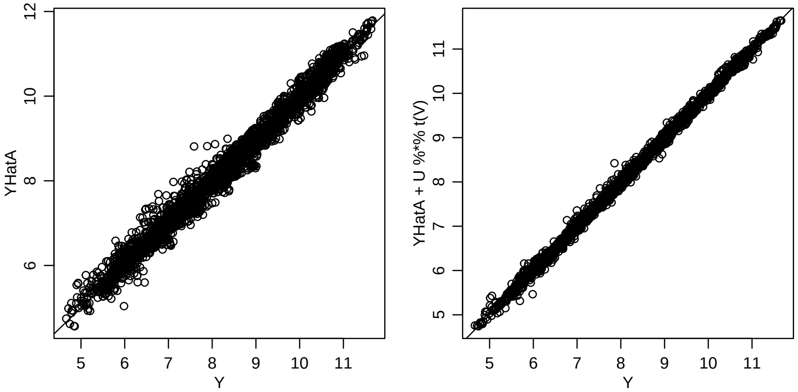
which is *linear* in the column effects $(b_j, \mathbf{v}_j)$.

Similarly,
Update row effects: Given column effects $(b_j, \mathbf{v}_j)$,

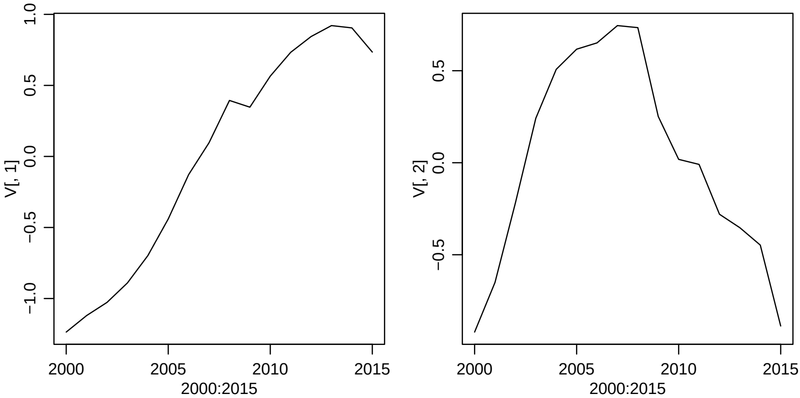
$$\begin{aligned} y_{i,j} &= \mu + (a_i, \mathbf{u}_i)^\top (b_j, \mathbf{v}_j) + \epsilon_{i,j} \\ &\equiv \mu + \beta_i^\top \mathbf{x}_j + \epsilon_{i,j} \end{aligned}$$

which is *linear* in the row effects (a_i, u_j).

Two-dimensional GDP model



Two-dimensional GDP model



Two-dimensional GDP model