Interactions in multilevel models

Andrew Gelman, Samantha Cook, and Shouhao Zhou
Department of Statistics and Department of Political Science
Columbia University

9 Aug 2005
Multilevel models and interactions

- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)
- Collaborators:
  - Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens
Multilevel models and interactions

- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)
- Collaborators:
  - Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens
Multilevel models and interactions

- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)
- Collaborators:
  - Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens
Multilevel models and interactions

- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)

Collaborators:
- Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens
Multilevel models and interactions

- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)
- Collaborators:
  - Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher’s classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i \theta + X_i \beta + \epsilon_i$

```
control
treatment
```

"before" measurement, x
"after" measurement, y

Andrew Gelman, Samantha Cook, and Shouhao Zhao

Interactions in multilevel models
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i \theta + X_i \beta + \epsilon_i$

![Scatter plot showing before-after measurement data with treatment and control groups.](image-url)
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher’s classical null hyp: effect is zero for all cases
  - Regression model: \( y_i = T_i \theta + X_i \beta + \epsilon_i \)
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher’s classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i \theta + X_i \beta + \epsilon_i$
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher’s classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i \theta + X_i \beta + \epsilon_i$
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for controls than for treated units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

Andrew Gelman, Samantha Cook, and Shouhao Zhao

Interactions in multilevel models
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for controls than for treated units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for controls than for treated units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for controls than for treated units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for controls than for treated units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats
Actual data show interactions

- Treatment interacts with “before” measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats
Observational study of legislative redistricting: before-after data
Educational experiment: correlation between pre-test and post-test data for controls and for treated units

![Graph showing correlation between grades and pre-test/post-test data for controls and treated units.](image-url)
Correlation between two successive Congressional elections for incumbents running (controls) and open seats (treated)
Interactions as variance components

Unit-level “error term” \( \eta_i \)

- For control units, \( \eta_i \) persists from time 1 to time 2
- For treatment units, \( \eta_i \) changes:
  - Subtractive treatment error (\( \eta_i \) only at time 1)
  - Additive treatment error (\( \eta_i \) only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” \( \eta_i \)

- For control units, \( \eta_i \) persists from time 1 to time 2
- For treatment units, \( \eta_i \) changes:
  - Subtractive treatment error (\( \eta_i \) only at time 1)
  - Additive treatment error (\( \eta_i \) only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” $\eta_i$

- For control units, $\eta_i$ persists from time 1 to time 2
- For treatment units, $\eta_i$ changes:
  - Subtractive treatment error ($\eta_i$ only at time 1)
  - Additive treatment error ($\eta_i$ only at time 2)
  - Replacement treatment error

- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” $\eta_i$

- For control units, $\eta_i$ persists from time 1 to time 2
- For treatment units, $\eta_i$ changes:
  - Subtractive treatment error ($\eta_i$ only at time 1)
  - Additive treatment error ($\eta_i$ only at time 2)
  - Replacement treatment error

- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” $\eta_i$

- For control units, $\eta_i$ persists from time 1 to time 2
- For treatment units, $\eta_i$ changes:
  - Subtractive treatment error ($\eta_i$ only at time 1)
  - Additive treatment error ($\eta_i$ only at time 2)
  - Replacement treatment error

- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” $\eta_i$

- For control units, $\eta_i$ persists from time 1 to time 2
- For treatment units, $\eta_i$ changes:
  - Subtractive treatment error ($\eta_i$ only at time 1)
  - Additive treatment error ($\eta_i$ only at time 2)
  - Replacement treatment error

- Under all these models, the before-after correlation is higher for controls than treated units
Interactions as variance components

Unit-level “error term” $\eta_i$

- For control units, $\eta_i$ persists from time 1 to time 2
- For treatment units, $\eta_i$ changes:
  - Subtractive treatment error ($\eta_i$ only at time 1)
  - Additive treatment error ($\eta_i$ only at time 2)
  - Replacement treatment error

- Under all these models, the before-after correlation is higher for controls than treated units
Summary of first part of talk

- Treatment interactions are important
- Before-after correlations are *lower* in treatment group
- Interpret as additional variance component that is altered by the treatment
Summary of first part of talk

- Treatment interactions are important
  - Before-after correlations are *lower* in treatment group
  - Interpret as additional variance component that is altered by the treatment
Summary of first part of talk

- Treatment interactions are important
- Before-after correlations are *lower* in treatment group
- Interpret as additional variance component that is altered by the treatment
Summary of first part of talk

- Treatment interactions are important
- Before-after correlations are lower in treatment group
- Interpret as additional variance component that is altered by the treatment
Examples of interactions in regression

- Federal spending by state, year, category ($50 \times 19 \times 10$)
- Vote preference given state and demographic variables ($50 \times 2 \times 2 \times 4 \times 4$)
- Rich state, poor state, red state, blue state ($50 \times 2$ for each election)
- Meta-analysis of incentives in sample surveys ($2^6$)
Examples of interactions in regression

- Federal spending by state, year, category (50 × 19 × 10)
- Vote preference given state and demographic variables (50 × 2 × 2 × 4 × 4)
- Rich state, poor state, red state, blue state (50 × 2 for each election)
- Meta-analysis of incentives in sample surveys (2^6)
Examples of interactions in regression

- Federal spending by state, year, category (50 × 19 × 10)
- Vote preference given state and demographic variables (50 × 2 × 2 × 4 × 4)
- Rich state, poor state, red state, blue state (50 × 2 for each election)
- Meta-analysis of incentives in sample surveys (2^6)
Examples of interactions in regression

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- Vote preference given state and demographic variables \((50 \times 2 \times 2 \times 4 \times 4)\)
- Rich state, poor state, red state, blue state \((50 \times 2\) for each election)\)
- Meta-analysis of incentives in sample surveys \((2^6)\)
Examples of interactions in regression

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- Vote preference given state and demographic variables \((50 \times 2 \times 2 \times 4 \times 4)\)
- Rich state, poor state, red state, blue state \((50 \times 2 \text{ for each election})\)
- Meta-analysis of incentives in sample surveys \((2^6)\)
General concerns

- Lots of potential interactions
- Setting high-level interactions to zero? Too extreme, especially when interactions are of substantive interest
- Simple hierarchical model for interactions is too crude
- Model: large main effects can have large interactions. In hierarchical setting, model should come “naturally”
General concerns

- Lots of potential interactions
- Setting high-level interactions to zero? Too extreme, especially when interactions are of substantive interest
- Simple hierarchical model for interactions is too crude
- Model: large main effects can have large interactions. In hierarchical setting, model should come “naturally”
General concerns

- Lots of potential interactions
- Setting high-level interactions to zero? Too extreme, especially when interactions are of substantive interest
- Simple hierarchical model for interactions is too crude
- Model: large main effects can have large interactions. In hierarchical setting, model should come “naturally”
General concerns

- Lots of potential interactions
- Setting high-level interactions to zero? Too extreme, especially when interactions are of substantive interest
- Simple hierarchical model for interactions is too crude
- Model: large main effects can have large interactions. In hierarchical setting, model should come “naturally”
General concerns

- Lots of potential interactions
- Setting high-level interactions to zero? Too extreme, especially when interactions are of substantive interest
- Simple hierarchical model for interactions is too crude
- Model: large main effects can have large interactions. In hierarchical setting, model should come “naturally”
Federal spending by state

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- For each spending category, \(50 \times 19\) data structure
- \(y_{jt} = \alpha_j + \beta_t + \gamma_{jt}\)
- Possible model: \(\gamma_{jt} \sim N(0, A + B |\alpha_j \beta_t|)\)
- Some example data
Federal spending by state

- Federal spending by state, year, category (50 × 19 × 10)
- For each spending category, 50 × 19 data structure
- \( y_{jt} = \alpha_j + \beta_t + \gamma_{jt} \)
- possible model: \( \gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|) \)
- Some example data
Federal spending by state

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- For each spending category, \(50 \times 19\) data structure
  - \(y_{jt} = \alpha_j + \beta_t + \gamma_{jt}\)
  - possible model: \(\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)\)
  - Some example data
Federal spending by state

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- For each spending category, \(50 \times 19\) data structure
- \(y_{jt} = \alpha_j + \beta_t + \gamma_{jt}\)
- possible model: \(\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)\)
- Some example data
Federal spending by state

- Federal spending by state, year, category (50 × 19 × 10)
- For each spending category, 50 × 19 data structure
- $y_{jt} = \alpha_j + \beta_t + \gamma_{jt}$
- Possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)$
- Some example data
Federal spending by state

- Federal spending by state, year, category \((50 \times 19 \times 10)\)
- For each spending category, \(50 \times 19\) data structure
- \(y_{jt} = \alpha_j + \beta_t + \gamma_{jt}\)
- possible model: \(\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)\)
- Some example data
Interactions in before-after studies
Interactions in regressions
Conclusions

Federal spending
Vote preferences
Income and voting
Incentives in sample surveys
Summary

Interactions $\gamma_{jt}$ plotted vs. main effects $|\alpha_j \beta_t|$
Logistic regression for pre-election polls

- Logistic regression: \( \Pr(y_i = 1) = \logit^{-1}((X\beta)_i) \)
- \( X \) includes demographic and geographic factors: sex, ethnicity, age, education, state
- Hierarchical model for 4 age levels, 4 education levels, 16 age \( \times \) education, 50 states
- Also consider interactions such as ethnicity \( \times \) state
Logistic regression for pre-election polls

- Logistic regression: \( \Pr(y_i = 1) = \logit^{-1}((X\beta)_i) \)
- \( X \) includes demographic and geographic factors: sex, ethnicity, age, education, state
- Hierarchical model for 4 age levels, 4 education levels, 16 age \( \times \) education, 50 states
- Also consider interactions such as ethnicity \( \times \) state
Logistic regression for pre-election polls

- Logistic regression: \( \Pr(y_i = 1) = \logit^{-1}((X\beta)_i) \)
- \( X \) includes demographic and geographic factors: sex, ethnicity, age, education, state
  - Hierarchical model for 4 age levels, 4 education levels, 16 age \( \times \) education, 50 states
  - Also consider interactions such as ethnicity \( \times \) state
Logistic regression for pre-election polls

- Logistic regression: \( \Pr(y_i = 1) = \logit^{-1}((X\beta)_i) \)
- \( X \) includes demographic and geographic factors: sex, ethnicity, age, education, state
- Hierarchical model for 4 age levels, 4 education levels, 16 age \( \times \) education, 50 states
- Also consider interactions such as ethnicity \( \times \) state
Logistic regression for pre-election polls

- Logistic regression: $\Pr(y_i = 1) = \logit^{-1}((X\beta)_i)$
- $X$ includes demographic and geographic factors: sex, ethnicity, age, education, state
- Hierarchical model for 4 age levels, 4 education levels, 16 age $\times$ education, 50 states
- Also consider interactions such as ethnicity $\times$ state
### Logistic regression with lots of predictors

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
<th>Rhat</th>
<th>n.eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.0</td>
<td>0.402</td>
<td>0.147</td>
<td>0.044</td>
<td>0.326</td>
<td>0.413</td>
<td>0.499</td>
<td>0.652</td>
<td>1.024</td>
<td>110</td>
</tr>
<tr>
<td>b.female</td>
<td>-0.094</td>
<td>0.102</td>
<td>-0.283</td>
<td>-0.162</td>
<td>-0.095</td>
<td>-0.034</td>
<td>0.107</td>
<td>1.001</td>
<td>1000</td>
</tr>
<tr>
<td>b.black</td>
<td>-1.701</td>
<td>0.305</td>
<td>-2.323</td>
<td>-1.910</td>
<td>-1.691</td>
<td>-1.486</td>
<td>-1.152</td>
<td>1.014</td>
<td>500</td>
</tr>
<tr>
<td>b.female.black</td>
<td>-0.143</td>
<td>0.393</td>
<td>-0.834</td>
<td>-0.383</td>
<td>-0.155</td>
<td>0.104</td>
<td>0.620</td>
<td>1.007</td>
<td>1000</td>
</tr>
<tr>
<td>B.age[1]</td>
<td>0.084</td>
<td>0.088</td>
<td>-0.053</td>
<td>0.012</td>
<td>0.075</td>
<td>0.140</td>
<td>0.277</td>
<td>1.062</td>
<td>45</td>
</tr>
<tr>
<td>B.age[2]</td>
<td>-0.072</td>
<td>0.087</td>
<td>-0.260</td>
<td>-0.121</td>
<td>-0.054</td>
<td>-0.006</td>
<td>0.052</td>
<td>1.017</td>
<td>190</td>
</tr>
<tr>
<td>B.age[3]</td>
<td>0.044</td>
<td>0.077</td>
<td>-0.105</td>
<td>-0.007</td>
<td>0.038</td>
<td>0.095</td>
<td>0.203</td>
<td>1.029</td>
<td>130</td>
</tr>
<tr>
<td>B.age[4]</td>
<td>-0.057</td>
<td>0.096</td>
<td>-0.265</td>
<td>-0.115</td>
<td>-0.052</td>
<td>0.001</td>
<td>0.133</td>
<td>1.076</td>
<td>32</td>
</tr>
<tr>
<td>B.edu[1]</td>
<td>-0.148</td>
<td>0.131</td>
<td>-0.436</td>
<td>-0.241</td>
<td>-0.137</td>
<td>-0.044</td>
<td>0.053</td>
<td>1.074</td>
<td>31</td>
</tr>
<tr>
<td>B.edu[2]</td>
<td>-0.022</td>
<td>0.082</td>
<td>-0.182</td>
<td>-0.069</td>
<td>-0.021</td>
<td>0.025</td>
<td>0.152</td>
<td>1.028</td>
<td>160</td>
</tr>
<tr>
<td>B.edu[3]</td>
<td>0.148</td>
<td>0.112</td>
<td>-0.032</td>
<td>0.065</td>
<td>0.142</td>
<td>0.228</td>
<td>0.370</td>
<td>1.049</td>
<td>45</td>
</tr>
<tr>
<td>B.edu[4]</td>
<td>0.023</td>
<td>0.090</td>
<td>-0.170</td>
<td>-0.030</td>
<td>0.015</td>
<td>0.074</td>
<td>0.224</td>
<td>1.061</td>
<td>37</td>
</tr>
<tr>
<td>B.age.edu[1,1]</td>
<td>-0.044</td>
<td>0.133</td>
<td>-0.363</td>
<td>-0.104</td>
<td>-0.019</td>
<td>0.025</td>
<td>0.170</td>
<td>1.018</td>
<td>1000</td>
</tr>
<tr>
<td>B.age.edu[1,2]</td>
<td>0.059</td>
<td>0.123</td>
<td>-0.153</td>
<td>-0.011</td>
<td>0.032</td>
<td>0.118</td>
<td>0.353</td>
<td>1.016</td>
<td>580</td>
</tr>
<tr>
<td>B.age.edu[1,3]</td>
<td>0.049</td>
<td>0.124</td>
<td>-0.146</td>
<td>-0.023</td>
<td>0.022</td>
<td>0.104</td>
<td>0.349</td>
<td>1.015</td>
<td>280</td>
</tr>
<tr>
<td>B.age.edu[1,4]</td>
<td>0.001</td>
<td>0.116</td>
<td>-0.237</td>
<td>-0.061</td>
<td>0.000</td>
<td>0.052</td>
<td>0.280</td>
<td>1.010</td>
<td>1000</td>
</tr>
<tr>
<td>B.age.edu[2,1]</td>
<td>0.066</td>
<td>0.152</td>
<td>-0.208</td>
<td>-0.008</td>
<td>0.032</td>
<td>0.124</td>
<td>0.449</td>
<td>1.022</td>
<td>140</td>
</tr>
<tr>
<td>B.age.edu[2,2]</td>
<td>-0.081</td>
<td>0.127</td>
<td>-0.407</td>
<td>-0.137</td>
<td>-0.055</td>
<td>0.001</td>
<td>0.094</td>
<td>1.057</td>
<td>120</td>
</tr>
<tr>
<td>B.age.edu[2,3]</td>
<td>-0.004</td>
<td>0.102</td>
<td>-0.226</td>
<td>-0.048</td>
<td>0.000</td>
<td>0.041</td>
<td>0.215</td>
<td>1.008</td>
<td>940</td>
</tr>
<tr>
<td>B.age.edu[2,4]</td>
<td>-0.042</td>
<td>0.108</td>
<td>-0.282</td>
<td>-0.100</td>
<td>-0.026</td>
<td>0.014</td>
<td>0.157</td>
<td>1.017</td>
<td>170</td>
</tr>
<tr>
<td>B.age.edu[3,1]</td>
<td>0.034</td>
<td>0.135</td>
<td>-0.215</td>
<td>-0.030</td>
<td>0.009</td>
<td>0.091</td>
<td>0.361</td>
<td>1.021</td>
<td>230</td>
</tr>
<tr>
<td>B.age.edu[3,2]</td>
<td>0.007</td>
<td>0.102</td>
<td>-0.213</td>
<td>-0.039</td>
<td>0.003</td>
<td>0.052</td>
<td>0.220</td>
<td>1.019</td>
<td>610</td>
</tr>
<tr>
<td>B.age.edu[3,3]</td>
<td>0.033</td>
<td>0.130</td>
<td>-0.215</td>
<td>-0.029</td>
<td>0.009</td>
<td>0.076</td>
<td>0.410</td>
<td>1.080</td>
<td>61</td>
</tr>
<tr>
<td>B.age.edu[3,4]</td>
<td>-0.009</td>
<td>0.109</td>
<td>-0.236</td>
<td>-0.064</td>
<td>-0.005</td>
<td>0.043</td>
<td>0.214</td>
<td>1.024</td>
<td>150</td>
</tr>
<tr>
<td>B.age.edu[4,1]</td>
<td>-0.141</td>
<td>0.190</td>
<td>-0.672</td>
<td>-0.224</td>
<td>-0.086</td>
<td>-0.003</td>
<td>0.100</td>
<td>1.036</td>
<td>270</td>
</tr>
<tr>
<td>B.age.edu[4,2]</td>
<td>-0.014</td>
<td>0.119</td>
<td>-0.280</td>
<td>-0.059</td>
<td>-0.008</td>
<td>0.033</td>
<td>0.239</td>
<td>1.017</td>
<td>240</td>
</tr>
</tbody>
</table>
### Bayesian Anova display

**Source**

- sex: 1
- ethnicity: 1
- sex * ethnicity: 1
- age: 3
- education: 3
- age * education: 9
- region: 3
- region * state: 46
- ethnicity * region: 3
- ethnicity * region * state: 46

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Est. sd of effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ethnicity</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>sex * ethnicity</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>education</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>age * education</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>region</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>region * state</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>ethnicity * region</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ethnicity * region * state</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>
Prediction error as function of # of predictors

MSE : training sample

MSE : test sample
Rich state, poor state, red state, blue state; or, What’s the matter with Connecticut?

- Richer voters favor the Republicans, but
- Richer states favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")
Rich state, poor state, red state, blue state; or, What’s the matter with Connecticut?

- Richer voters favor the Republicans, but
- Richer states favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")
Rich state, poor state, red state, blue state; or, What’s the matter with Connecticut?

- Richer voters favor the Republicans, *but*
- Richer states favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state ("varying-intercept model")
Rich state, poor state, red state, blue state; or, What’s the matter with Connecticut?

- Richer *voters* favor the Republicans, *but*
- Richer *states* favor the Democrats
- Hierarchical logistic regression: predict your vote given your income and your state (“varying-intercept model”)

Andrew Gelman, Samantha Cook, and Shouhao Zhao
Varying-intercept model

Andrew Gelman, Samantha Cook, and Shouhao Zhao
Varying-intercept, varying-slope model

Andrew Gelman, Samantha Cook, and Shouhao Zhao

Interactions in multilevel models
Interactions in multilevel models

Andrew Gelman, Samantha Cook, and Shouhao Zhao

Interactions in before-after studies
Interactions in regressions
Conclusions
Federal spending
Vote preferences
Income and voting
Incentives in sample surveys
Summary

Interactions!

Avg Income 2000 vs. Var Slope 2000

Avg State Income ($10k)
Slope
2.0 2.5 3.0 3.5
0.0 0.1 0.2 0.3 0.4

MS
AR
WV
LA
AL
NM
MT
KY
OK
SD
ND
MN
MS
MO
MT
NE
NV
NH
NJ
NM
NY
NC
ND
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
WA
WV
WI
WY
CT

Andrew Gelman, Samantha Cook, and Shouhao Zhao
Interactions in multilevel models
3-way interactions!

Andrew Gelman, Samantha Cook, and Shouhao Zhao

Interactions in before-after studies
Interactions in regressions
Conclusions
Federal spending
Vote preferences
Income and voting
Incentives in sample surveys
Summary

Interactions in multilevel models
Meta-analysis of effects of incentives on survey response rates

6 factors
- Incentive or not
- Value of incentive
- Form (gift or cash)
- Timing (before or after)
- Mode (telephone or face-to-face)
- Burden (short or long survey)

Models
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models

Andrew Gelman, Samantha Cook, and Shouhao Zhao
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
    - Mode (telephone or face-to-face)
    - Burden (short or long survey)
- Models
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
  - No interactions: estimates don’t make sense
  - Interactions: estimates are out of control
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
  - No interactions: estimates don’t make sense
  - Interactions: estimates are out of control
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
  - No interactions: estimates don't make sense
  - Interactions: estimates are out of control
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
  - No interactions: estimates don't make sense
  - Interactions: estimates are out of control
Meta-analysis of effects of incentives on survey response rates

- 6 factors
  - Incentive or not
  - Value of incentive
  - Form (gift or cash)
  - Timing (before or after)
  - Mode (telephone or face-to-face)
  - Burden (short or long survey)

- Models
  - No interactions: estimates don't make sense
  - Interactions: estimates are out of control
Model without interactions

<table>
<thead>
<tr>
<th></th>
<th>Beta (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.4 (1.6)</td>
</tr>
<tr>
<td>Value of incentive</td>
<td>0.34 (0.17)</td>
</tr>
<tr>
<td>Prepayment</td>
<td>2.8 (1.8)</td>
</tr>
<tr>
<td>Gift</td>
<td>-6.9 (1.5)</td>
</tr>
<tr>
<td>Burden</td>
<td>3.3 (1.3)</td>
</tr>
</tbody>
</table>

Will a low-value postpaid gift really *reduce* response rates by 7 percentage points??
Model without interactions

- Estimated effects on response rate (in percentage points)

<table>
<thead>
<tr>
<th></th>
<th>Beta (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.4 (1.6)</td>
</tr>
<tr>
<td>Value of incentive</td>
<td>0.34 (0.17)</td>
</tr>
<tr>
<td>Prepayment</td>
<td>2.8 (1.8)</td>
</tr>
<tr>
<td>Gift</td>
<td>-6.9 (1.5)</td>
</tr>
<tr>
<td>Burden</td>
<td>3.3 (1.3)</td>
</tr>
</tbody>
</table>

- Will a low-value postpaid gift really *reduce* response rates by 7 percentage points??
Model without interactions

- Estimated effects on response rate (in percentage points)

<table>
<thead>
<tr>
<th></th>
<th>Beta (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.4 (1.6)</td>
</tr>
<tr>
<td>Value of incentive</td>
<td>0.34 (0.17)</td>
</tr>
<tr>
<td>Prepayment</td>
<td>2.8 (1.8)</td>
</tr>
<tr>
<td>Gift</td>
<td>−6.9 (1.5)</td>
</tr>
<tr>
<td>Burden</td>
<td>3.3 (1.3)</td>
</tr>
</tbody>
</table>

- Will a low-value postpaid gift really reduce response rates by 7 percentage points??
Models with interactions

<table>
<thead>
<tr>
<th></th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>60.7 (2.2)</td>
<td>60.8 (2.5)</td>
<td>61.0 (2.5)</td>
<td>60.1 (2.5)</td>
</tr>
<tr>
<td>Incentive</td>
<td>5.4 (0.7 )</td>
<td>3.7 (0.8 )</td>
<td>2.8 (1.0 )</td>
<td>6.1 (1.2 )</td>
</tr>
<tr>
<td>Mode</td>
<td>15.2 (4.7)</td>
<td>16.1 (5.1)</td>
<td>16.0 (4.9)</td>
<td>18.0 (4.6)</td>
</tr>
<tr>
<td>Burden</td>
<td>−7.2 (4.3)</td>
<td>−8.9 (5.0)</td>
<td>−8.7 (5.0)</td>
<td>−9.9 (5.0)</td>
</tr>
<tr>
<td>Mode × Burden</td>
<td>−7.6 (9.8)</td>
<td>−7.8 (9.4)</td>
<td>−4.9 (9.1)</td>
<td></td>
</tr>
<tr>
<td>Incentive × Value</td>
<td>0.14 (0.03)</td>
<td>0.33 (0.09)</td>
<td>0.26 (0.09)</td>
<td></td>
</tr>
<tr>
<td>Incentive × Timing</td>
<td>4.4 (1.3 )</td>
<td>1.7 (1.7 )</td>
<td>−0.2 (2.1 )</td>
<td></td>
</tr>
<tr>
<td>Incentive × Form</td>
<td>1.4 (1.3 )</td>
<td>1.1 (1.2 )</td>
<td>−1.2 (2.0 )</td>
<td></td>
</tr>
<tr>
<td>Incentive × Mode</td>
<td>−2.3 (1.6)</td>
<td>−2.0 (1.7)</td>
<td>7.8 (2.9 )</td>
<td></td>
</tr>
<tr>
<td>Incentive × Burden</td>
<td>4.8 (1.5)</td>
<td>5.4 (1.8 )</td>
<td>−5.2 (2.7 )</td>
<td></td>
</tr>
<tr>
<td>Incentive × Value × Timing</td>
<td>0.40 (0.17)</td>
<td>0.58 (0.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Value × Burden</td>
<td>−0.06 (0.06)</td>
<td>1.10 (0.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Timing × Burden</td>
<td></td>
<td>11.1 (3.9 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Value × Form</td>
<td></td>
<td>0.30 (0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Value × Mode</td>
<td></td>
<td>−1.20 (0.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Timing × Form</td>
<td></td>
<td>9.9 (2.7 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Timing × Mode</td>
<td></td>
<td>−17.4 (4.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Form × Mode</td>
<td></td>
<td>−0.3 (2.5 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Form × Burden</td>
<td></td>
<td>5.9 (3.2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive × Mode × Burden</td>
<td></td>
<td>−5.8 (3.0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-study sd, $\sigma$</td>
<td>4.2 (0.3)</td>
<td>3.6 (0.3)</td>
<td>3.6 (0.3)</td>
<td>2.8 (0.3)</td>
</tr>
<tr>
<td>Between-study sd, $\tau$</td>
<td>18 (2)</td>
<td>19 (2)</td>
<td>18 (2)</td>
<td>18 (2)</td>
</tr>
</tbody>
</table>
Summary of second part of talk

- With many predictors come many many potential interactions
- Interactions can be crucial to substantive understanding
- Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters
- Simple inclusion of additional batches of interactions can hurt predictive power
- Goal: models where large main effects are more likely to have large interactions
- possible model: \( \gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|) \)
- But we really don't know yet what will work!
With many predictors come many many potential interactions

Interactions can be crucial to substantive understanding

Simple pooling of high-level interactions (“Anova” or even “Bayesian Anova”) is too crude, does not respect the structure of the parameters

Simple inclusion of additional batches of interactions can hurt predictive power

Goal: models where large main effects are more likely to have large interactions

possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)$

But we really don’t know yet what will work!
With many predictors come many many potential interactions

Interactions can be crucial to substantive understanding

Simple pooling of high-level interactions (“Anova” or even “Bayesian Anova”) is too crude, does not respect the structure of the parameters

Simple inclusion of additional batches of interactions can hurt predictive power

Goal: models where large main effects are more likely to have large interactions

possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)$

But we really don’t know yet what will work!
Summary of second part of talk

- With many predictors come many many potential interactions
- Interactions can be crucial to substantive understanding
- Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters
- Simple inclusion of additional batches of interactions can hurt predictive power
- Goal: models where large main effects are more likely to have large interactions
- Possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)$
- But we really don’t know yet what will work!
With many predictors come many many potential interactions

Interactions can be crucial to substantive understanding

Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters

Simple inclusion of additional batches of interactions can hurt predictive power

Goal: models where large main effects are more likely to have large interactions

Possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|)$

But we really don’t know yet what will work!
Summary of second part of talk

- With many predictors come many many potential interactions
- Interactions can be crucial to substantive understanding
- Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters
- Simple inclusion of additional batches of interactions can hurt predictive power
- Goal: models where large main effects are more likely to have large interactions
  - possible model: $\gamma_{jt} \sim N(0, A + B|\alpha_j \beta_t|)$
  - But we really don’t know yet what will work!
Summary of second part of talk

- With many predictors come many many potential interactions
- Interactions can be crucial to substantive understanding
- Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters
- Simple inclusion of additional batches of interactions can hurt predictive power
- Goal: models where large main effects are more likely to have large interactions
- Possible model: \( \gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|) \)
- But we really don't know yet what will work!
Summary of second part of talk

- With many predictors come many many potential interactions
- Interactions can be crucial to substantive understanding
- Simple pooling of high-level interactions ("Anova" or even "Bayesian Anova") is too crude, does not respect the structure of the parameters
- Simple inclusion of additional batches of interactions can hurt predictive power
- Goal: models where large main effects are more likely to have large interactions
- Possible model: \( \gamma_{jt} \sim N(0, A + B|\alpha_j\beta_t|) \)
- But we really don’t know yet what will work!
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, ...
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, …
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, etc.
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, . . .
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, . . .
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
Structured hierarchical models

- Need to go beyond exchangeability to shrink batches of parameters in a reasonable way
- For example, parameter matrices $\alpha_{jk}$ don’t look like exchangeable vectors
- Similar problems arise in shrinking higher-order terms in neural nets, wavelets, tree models, image models, . . .
- Recall the “blessing of dimensionality”: as the number of factors, and the number of levels per factor, increases, more information is available to estimate the hyperparameters of the big model
- In the background: advances in Bayesian computation including parameter expansion (Meng, Liu, Liu, Rubin, van Dyk), adaptive Metropolis algorithms (Pasarica), structured computations (Kerman)
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, ... interactions in regression models
- Appropriate models have lots of structure
- We need to try out different classes of models and see what works
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, \ldots, interactions in regression models
- Appropriate models have lots of structure
- We need to try out different classes of models and see what works
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Appropriate models have lots of structure
- We need to try out different classes of models and see what works
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, ..., interactions in regression models
- Appropriate models have lots of structure
- We need to try out different classes of models and see what works
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, . . . , interactions in regression models
- Appropriate models have lots of structure
  - We need to try out different classes of models and see what works
What have we learned?

- Interactions are important
  - Treatment interactions in before-after studies
  - 2-way, 3-way, ..., interactions in regression models
- Appropriate models have lots of structure
- We need to try out different classes of models and see what works