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

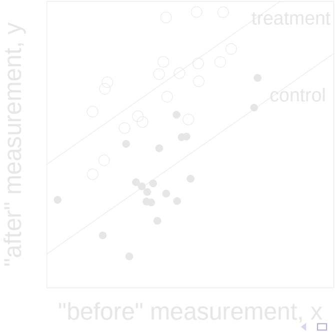
- Interactions in before-after studies
- Interactions in regressions with many input variables
- Many questions, few answers (yet)
- Collaborators:
  - Jouni Kerman, Iain Pardoe, Boris Shor, David Park, Joe Bafumi, Gary King, Zaiying Huang, Valerie Chan, Matt Stevens

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: $y_i = T_i\theta + X_i\beta + \epsilon_i$

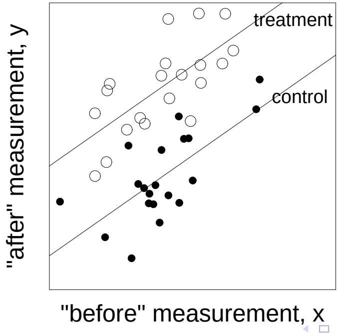


No-interaction model

- 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: $y_i = T_i\theta + X_i\beta + \epsilon_i$

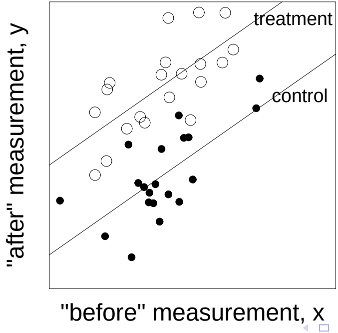
No-interaction model

- 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: $y_i = T_i\theta + X_i\beta + \epsilon_i$



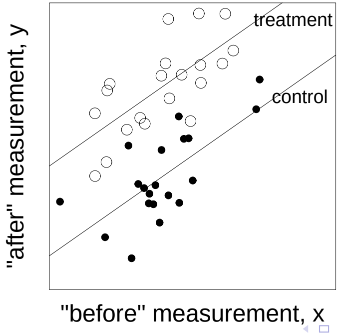
No-interaction model

- 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: $y_i = T_i\theta + X_i\beta + \epsilon_i$



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

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

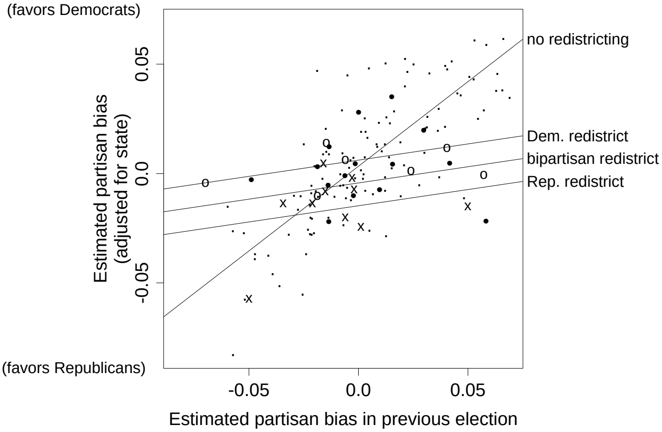
Actual data show interactions

- 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

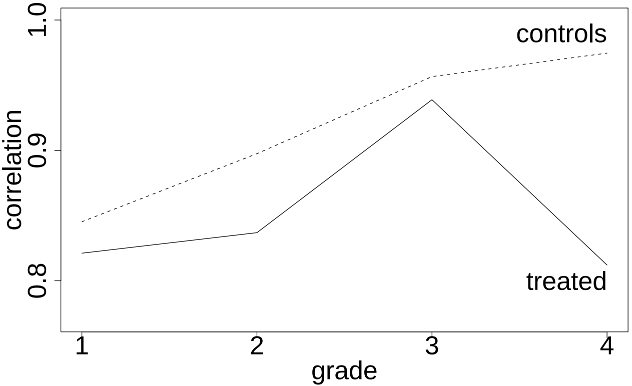
Actual data show interactions

- 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

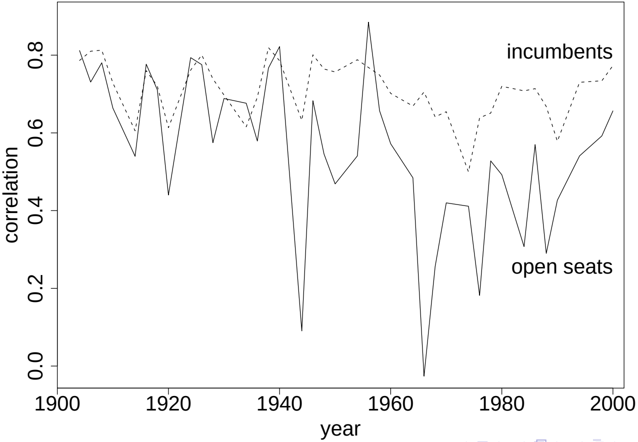
Observational study of legislative redistricting: before-after data



Educational experiment: correlation between pre-test and post-test data for controls and for treated units



Correlation between two successive Congressional elections for incumbents running (controls) and open seats (treated)



Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_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" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
    - Subtractive treatment error (η_i only at time 1)
    - Additive treatment error (η_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 (50×2 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 (50×2 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 (50×2 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 (50×2 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 (50×2 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, 50×19 data structure
- $y_{jt} = \alpha_j + \beta_t + \gamma_{jt}$
- possible model: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- Some example data

Federal spending by state

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

Federal spending by state

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

Federal spending by state

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

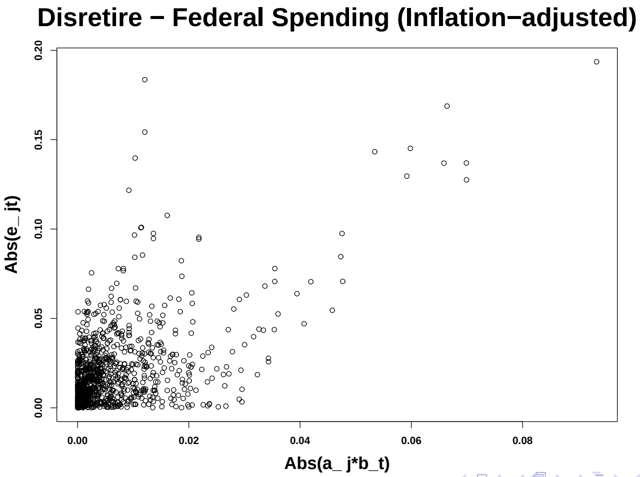
Federal spending by state

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

Federal spending by state

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

Interactions $|\gamma_{jt}$ plotted vs. main effects $|\alpha_j\beta_t|$



Logistic regression for pre-election polls

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- 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: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- 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: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- 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: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- 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

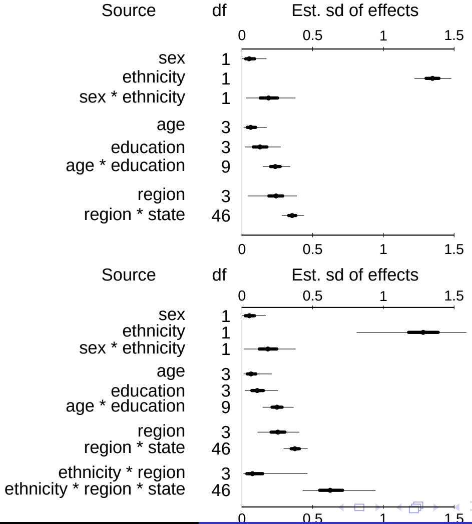
Logistic regression for pre-election polls

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- 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

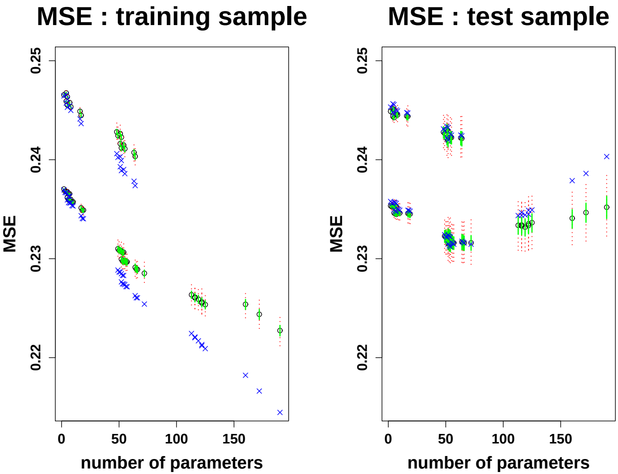
Logistic regression with lots of predictors

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

Bayesian Anova display



Prediction error as function of # of predictors



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")

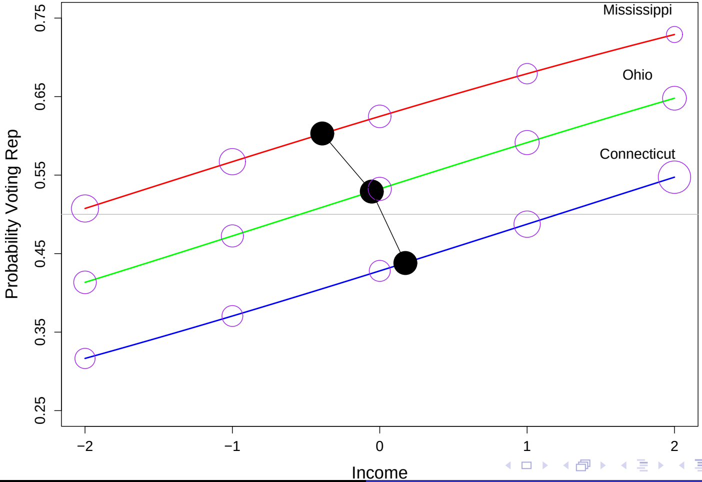
Rich state, poor state, red state, blue state; or, What's the matter with Connecticut?

- 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")

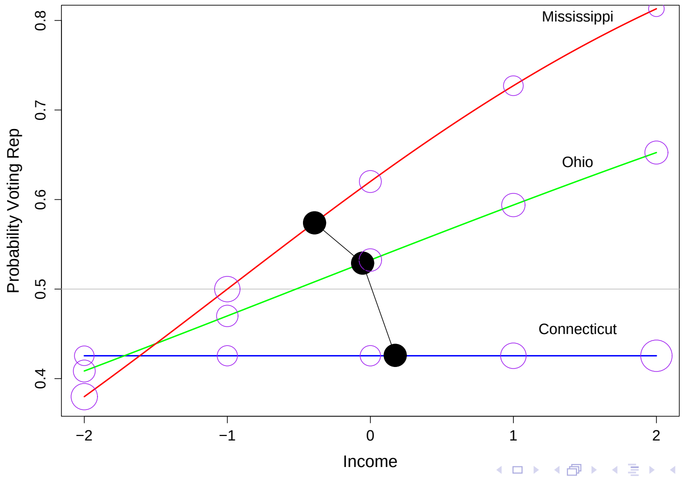
Rich state, poor state, red state, blue state; or, What's the matter with Connecticut?

- 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")

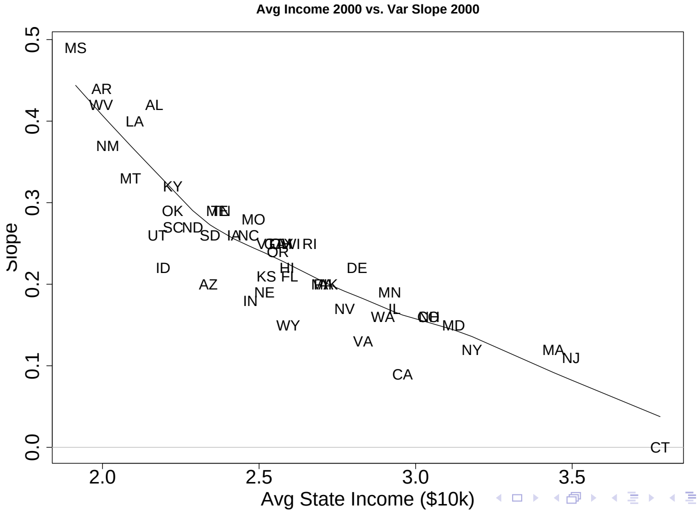
Varying-intercept model



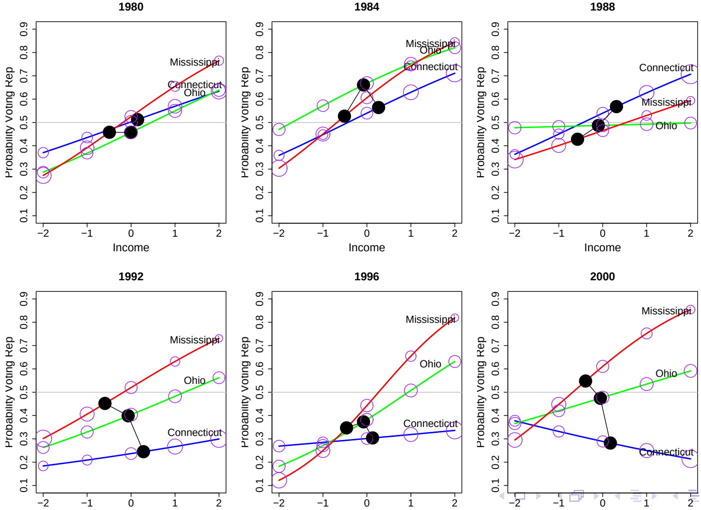
Varying-intercept, varying-slope model



Interactions!



3-way interactions!



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

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

- Estimated effects on response rate (in percentage points)

	Beta (s.e.)
Intercept	1.4 (1.6)
Value of incentive	0.34 (0.17)
Prepayment	2.8 (1.8)
Gift	−6.9 (1.5)
Burden	3.3 (1.3)

- 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)

	Beta (s.e.)
Intercept	1.4 (1.6)
Value of incentive	0.34 (0.17)
Prepayment	2.8 (1.8)
Gift	−6.9 (1.5)
Burden	3.3 (1.3)

- 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)

	Beta (s.e.)
Intercept	1.4 (1.6)
Value of incentive	0.34 (0.17)
Prepayment	2.8 (1.8)
Gift	−6.9 (1.5)
Burden	3.3 (1.3)

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

Models with interactions

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

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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0,\, A + B|\alpha_j \beta_t|)$
- 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: $\gamma_{jt} \sim \mathrm{N}(0, A + B|\alpha_j \beta_t|)$
- 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* α_{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* α_{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* α_{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* α_{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* α_{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* α_{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, . . . , 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