

Design and Analysis of Sample Surveys

Andrew Gelman

Department of Statistics and Department of Political Science
Columbia University

Class 12b: Challenges in multilevel regression and
poststratification

Choices in setting up MRP

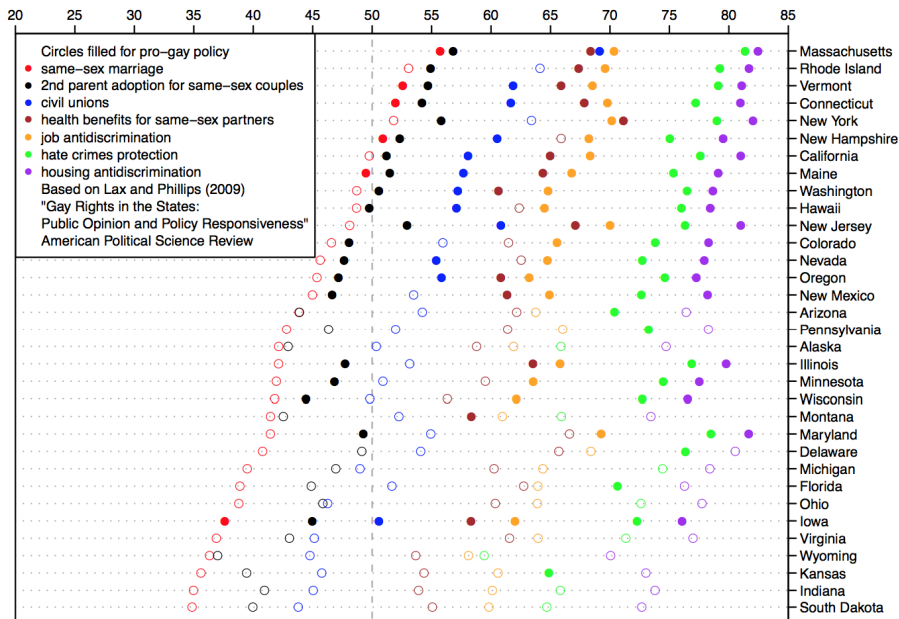
- ▶ Multilevel regression
 - ▶ How many and which factors to include?
 - ▶ Main effects and interactions
 - ▶ Individual-level predictors
 - ▶ Group-level predictors
- ▶ Poststratification
 - ▶ Census, CPS, ACS, PES
 - ▶ Poststratifying on party ID and on vote choice

- ▶ R package `mrp`
- ▶ Multilevel regression using R package `blme`, poststratifying by programming in R
- ▶ Challenge of putting together the `mrp` package
- ▶ Stata, etc.?

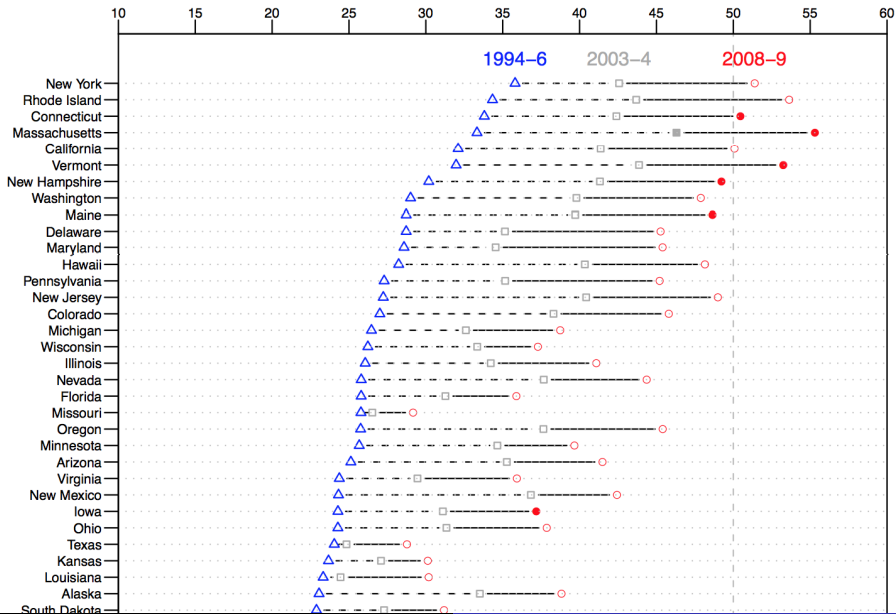
Multistage MRP

- ▶ Turnout and vote choice
- ▶ Multiple options

MRP with multiple questions: Gay rights



Time trends



MRP and ideal-point modeling

- ▶ Lax and Phillips looked at 39 policies:
- ▶ *Abortion*: require counseling, parental consent, parental notification, late-term abortions, waiting period
- ▶ *Education*: affirmative action, charter schools, standardized testing, vouchers
- ▶ *Electoral reform*: corporate/union campaign contributions, individual contributions, photo ID, recall elections, term limits
- ▶ *Gambling*: casinos, state lottery
- ▶ *Gay rights*: adoption, civil unions, health insurance, marriage, sodomy, employment nondiscrimination, hate crimes, housing nondiscrimination
- ▶ *Health care*: physician-assisted suicide, Medicaid availability, medical marijuana, children's health insurance, stem cell research
- ▶ *Immigration*: bilingual education, driver's licences for illegal immigrants, children of illegals in state colleges, verifying citizenship in hiring

Interrogating an MRP fit

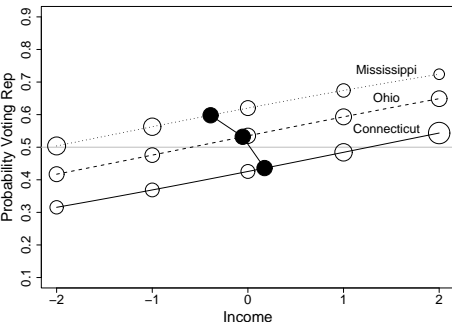
- ▶ Utah
- ▶ Vermont

Red state, blue state, rich state, poor state

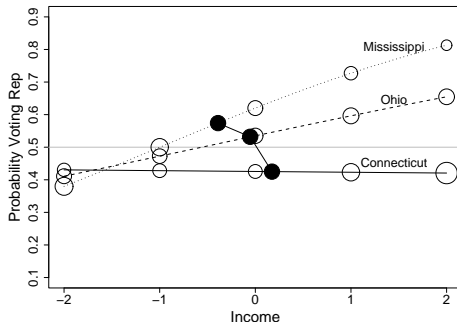
- ▶ Richer *voters* favor the Republicans, *but*
- ▶ Richer *states* favor the Democrats
- ▶ Hierarchical logistic regression: predict your vote given your income and your state (“varying-intercept model”)

Varying-intercept model, then model criticism, then varying-slope model

Varying-intercept model, 2000

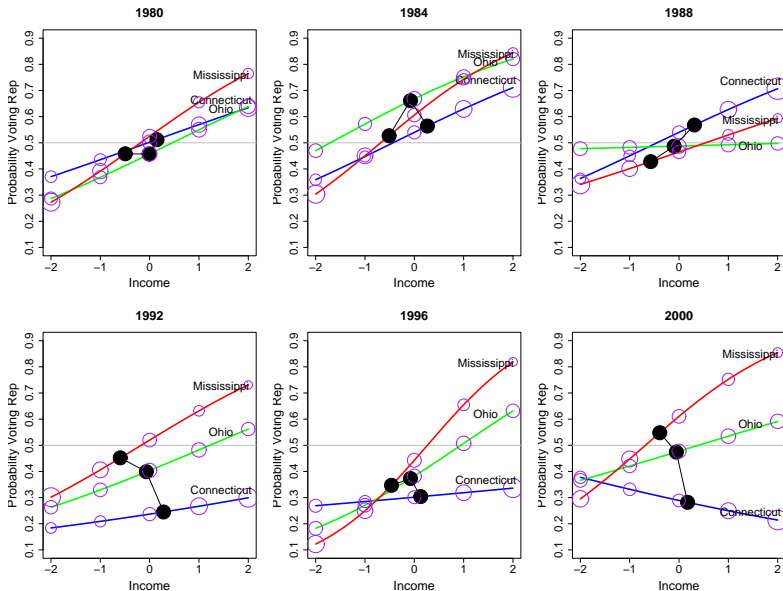


Varying-intercept, varying-slope model, 2000



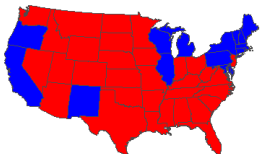
In any given state, the estimates would not be statistically significant!

3-way interactions!

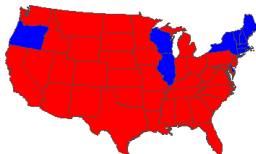


Adding another factor: The inference ...

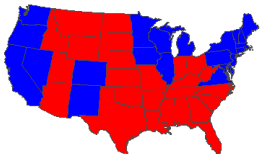
State winners in 2008 (rich voters only)



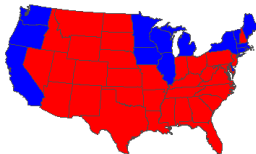
State winners in 2008 (rich Whites only)



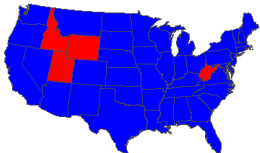
State winners in 2008 (middle-income voters)



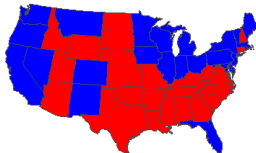
State winners in 2008 (middle-income Whites)



State winners in 2008 (poor voters only)



State winners in 2008 (poor Whites only)



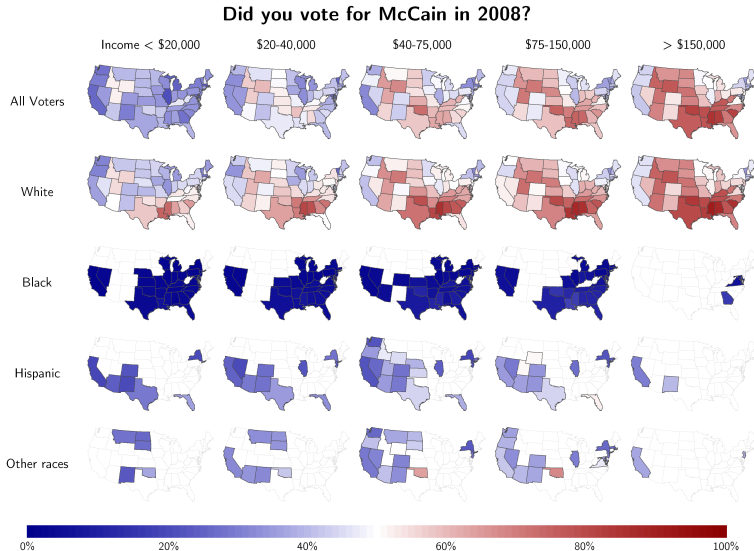
...and the refutation!

- ▶ Criticisms from the blogger “Daily Kos”:
 - ▶ Criticisms of the inferences:

“While Gelman claims only the under-\$20K white demo went for Obama, the results were far different. Per the exit poll — real voters — Obama won all whites: 54–45 percent for those making under \$50K, and 51–47% for those making over \$50K. ... New Hampshire is solidly Blue unlike Gelman’s maps, 58–40 — one of the most obvious misses in Gelman’s analysis. ...”
 - ▶ Criticisms of the method:

“Gelman inexplicably avoids using exit poll data ... while exit polls have their own margin of errors and sample composition problems, they sure as heck beat anything done over the telephone.”
 - ▶ Traditional statistical “conservatism” will be no defense here!

After improving the model



When a category represents less than 1% of the voters in a state, the state is left blank

A graph we made to study and criticize our inferences

2008 election: McCain share of the two-party vote in each income category within each state among all voters (gray) and just non-Hispanic whites (orange)

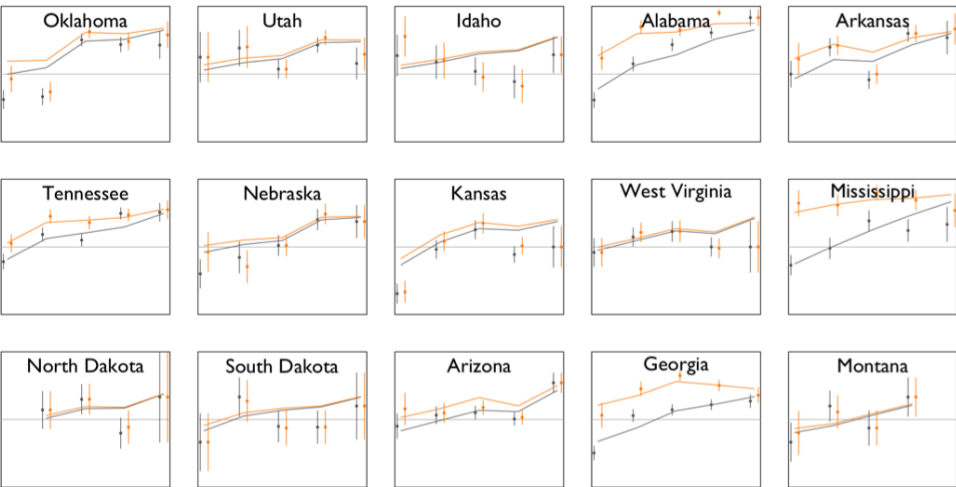


Dots are weighted averages from pooled June-Nov Pew surveys; error bars show ± 1 s.e. bounds.

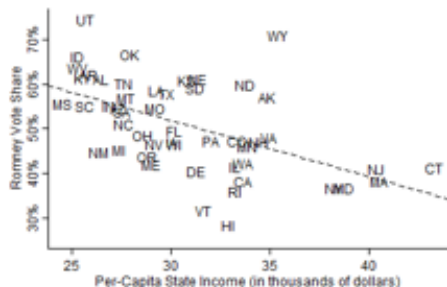
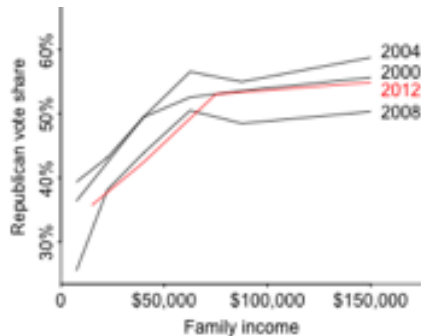
Curves are estimated using multilevel models and have a s.e. of about 3% at each point.

In detail ...

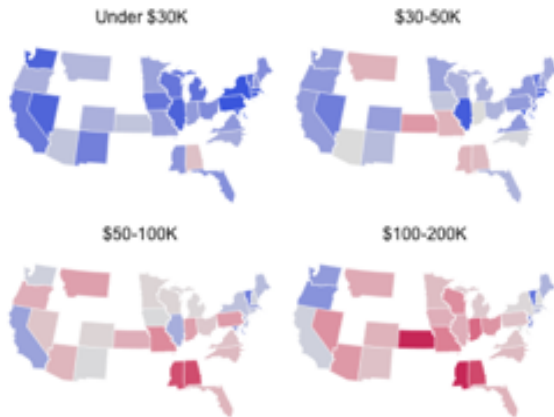
2008 election: McCain share of the two-party vote in each income category within each state among all voters (gray) and just non-Hispanic whites (orange)



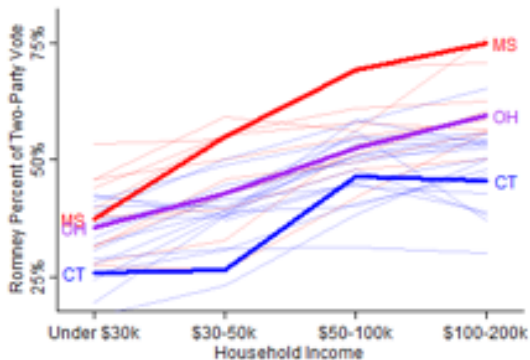
Rich and poor in the 2012 election



Obama vs. Romney by income and state



Varying slopes by state

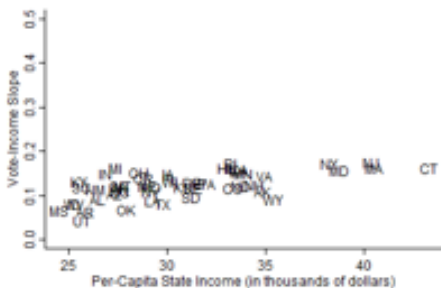


“Red State, Blue State” is dead

Exit Poll Results: Unpooled Regression Estimates



Shor-Moore Survey Results: Multilevel Regression Estimates



Open questions in MRP

- ▶ Deep interactions
- ▶ Poststratifying on non-census variables
 - ▶ Example: religious attendance
 - ▶ Example: party id
- ▶ Partial information