

# Define Ecological Inference problem

1. **cross-level inferences**
2. **From aggregate level to individual level**
3. **Data are collected at the aggregate (macro) level (e.g., 53% of the votes cast in the nation went to Obama in 2008; we also know Obama's state-level vote distribution)**
4. **We have no complete knowledge of aggregation process (i.e., how individuals voted).**
5. **A problem called “underidentification” (see Liu, 2007)**

# solutions

- Add assumptions about data aggregation process
- Statistical methods
  - single regression*
  - double regression*
  - neighborhood method*
  - King's EI Basic Model*
  - King's EI Extended Model*

# Racially polarized voting and voting-rights litigations

- The common goal statistically:

*to find whether there is a high-level of racial polarization for the black candidate in a biracial election*

- Example: Barack Obama in the 2008 presidential election. Can you define an ecological inference problem for this election?

# Obama in 2008

- First, let's see Obama's racial vote at the national level. According to the exit poll at the national level, Obama received votes from:
  - 43% of whites
  - 95% of blacks
  - 67% of Asians
  - 62% of Latinos
- From national level to state level, again, let's see the exit poll at the state level (see Liu 2010, Table 7.1)

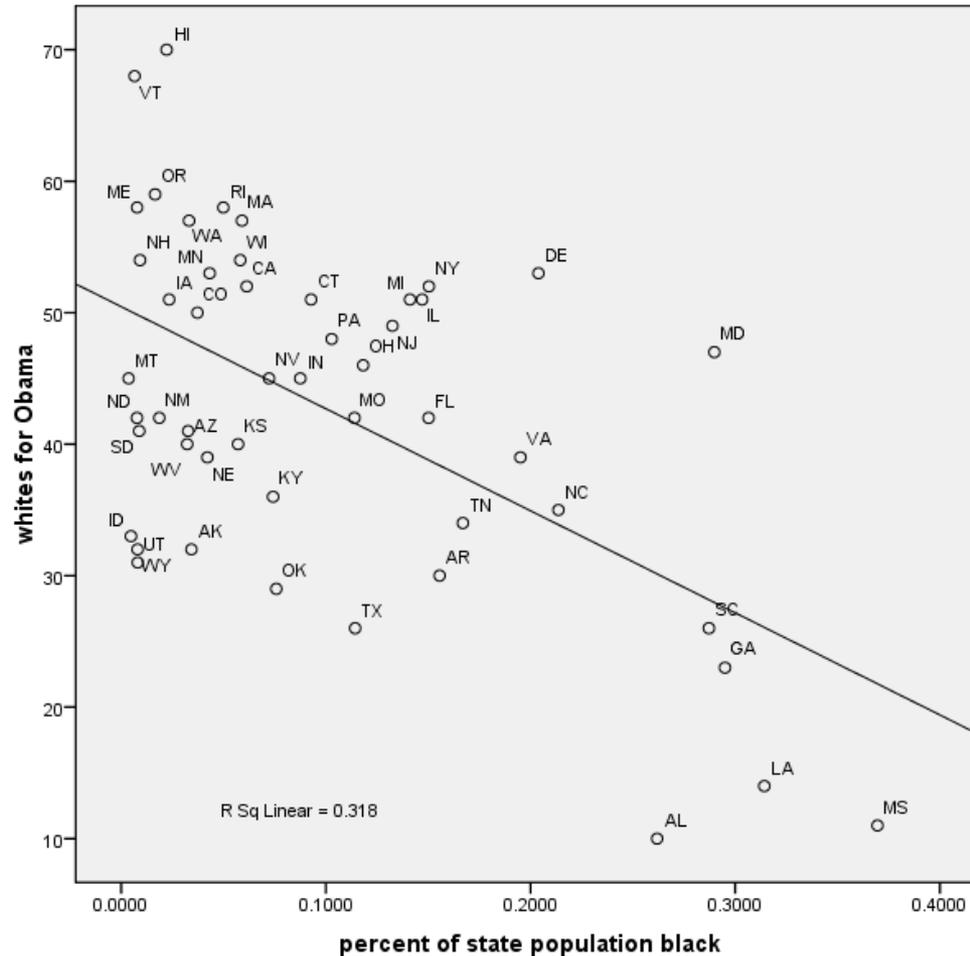
# What can you conclude?

- State-level white votes for Obama varied in different states. In other words, what we observed at the national level did not happen at the state level, which is a typical ecological inference problem----the fear of ecological fallacy.
- This is a major reason why King's EI method is better than previous methods, because it allows researchers to assume state-level differences in white votes for Obama (see Liu 2007, Table 1).
- Regression-based ecological inferences do not allow this, because they assume the quantities of interest are constant across ecological units

# What caused the differences in state-level white votes for Obama?

- Ecological inferences are both a science and an art
- You need to have a scientific method that can be applied to the right datasets
- You also need to have a creative mind to come up with a theory to explain the variance in your dependent variable
- Let's take a look at one possible solution

# A typical explanation for state-level white votes for Obama: black density



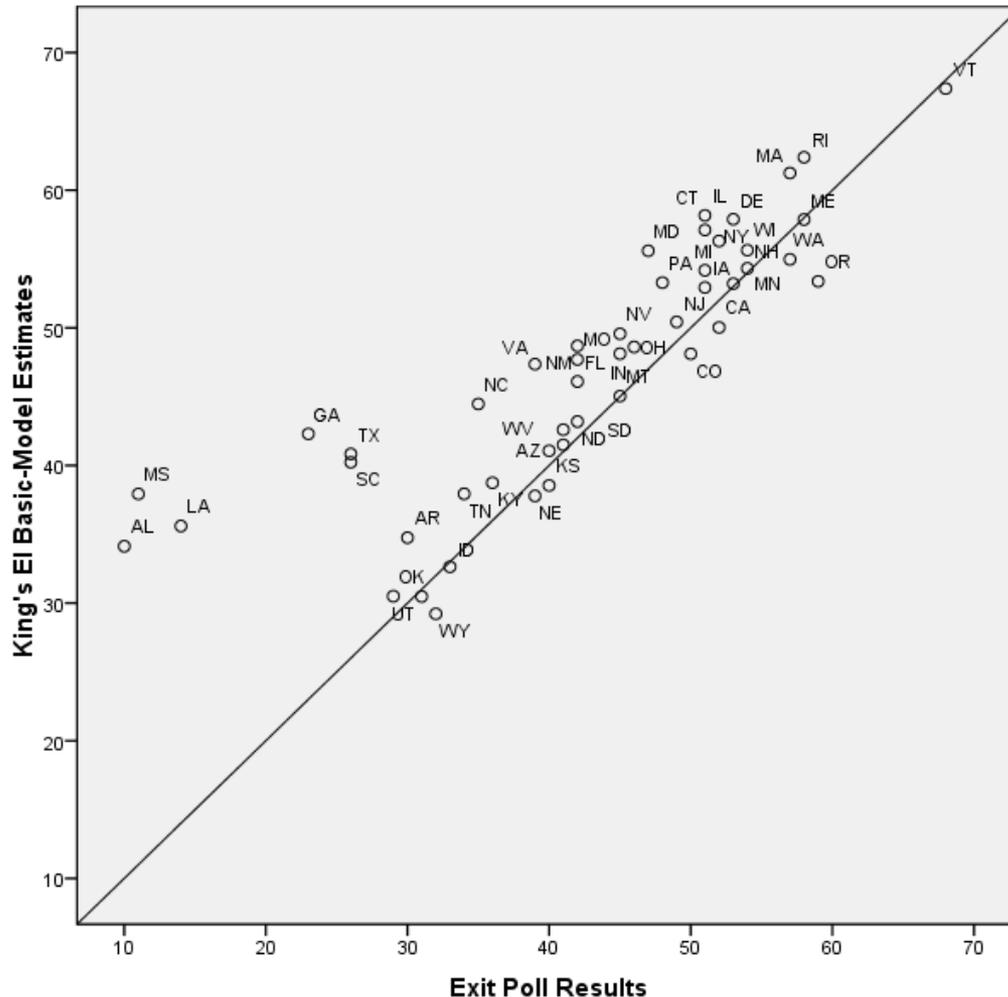
# Why black density?

- V. O. Key's famous black threat hypothesis: individuals respond to racial environmental variable
- Black density as a variable is easy to collect data
- All major ecological inference methods use black density as the main independent variable (see Liu 2007)

# Let's take a look at King's Ezi basic version

- Download Ezi from Gary King's web site
- Easy installation
- Let's use sample.fmt file from King's data set in Ezi software
- Go to "Model" then "specify"
- The x variable is the black density variable
- Note that the x variable is collected at the lowest level possible, King's Ezi allows researchers to compute racial voting (the quantity of interest) at both this lowest level and one-level above x variable.
- e.g., from state to the nation, from county to the nation, or from precinct to electoral district---you should always know your unit of analysis
- Hit "Run"
- After Ezi completes the operation, go to "read" to get the output. Let's use "Paggs" to find the macro-level estimates in the particular election (both racial voting and standard errors); "beta" for the lowest-level estimates

# Compare EI basic model with exit poll: the Case of 2008 Election



# Results show the limit on EI Basic model

- If the EI basic estimates of white support for Obama were completely converged to the exit poll results, then we should see all states clustered onto the 45-degree diagonal.
- EI basic model, in particular, estimated that the white support for Obama in the five Deep South states (AL, MS, LA, GA, SC) around 40% level. These estimates clearly overestimated his white vote share in the Deep South, whereas exit polls appear to provide more accurate results of white vote for Obama between 10% and 25% levels.
- Again, EI basic only include the black density as an independent variable, and  $n$  as a weight variable to give appropriate and different influences based on the electoral district sizes.

# The limits of black density

- Black density may correlate with other factors. e.g., look back at white vote for Obama and black density at the state level, the negative relationship may be caused by other variables, such as political culture (Elazar 1984), social capital (Putnam 2000), and racial tension (Liu 2010).
- EI basic version assumes that there should not be a correlation between black density and Obama's white vote in the first place, i.e., no contextual effect, or aggregation bias (see Liu 2007 for EI's other two assumptions).
- EI basic model can be misleading, if contextual effect exists (see Liu 2007 for a case study). EI basic may perform worse than regression-based ecological inference methods.

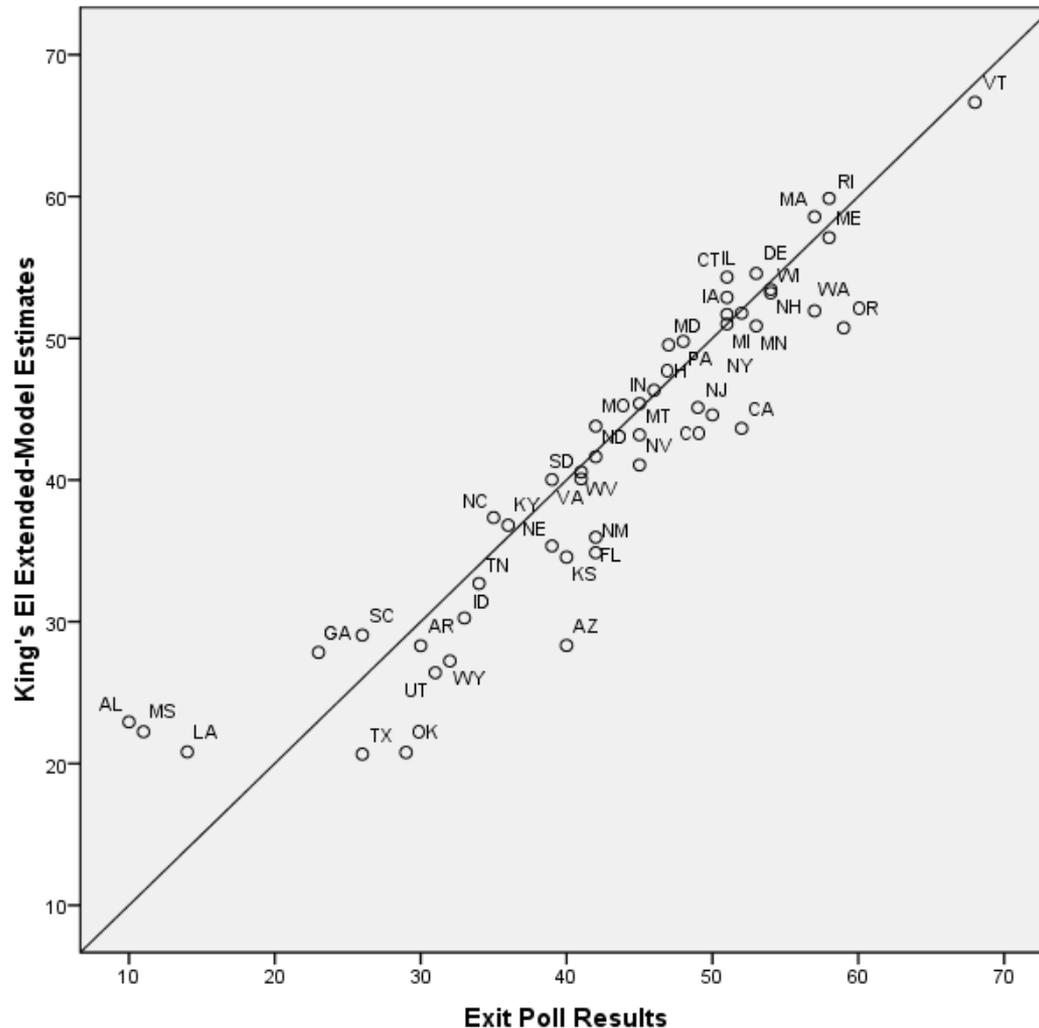
# What if the aggregation bias indeed exists?

- The only method available right now is to use King's EI Extended Model
- EI Extended Model allows contextual effect, i.e., the correlation between the quantity of interest and black density
- EI Extended Model requires one additional variable, a covariate that can help explain the variance in the dependent variable (i.e., racial vote for Obama in our example)

Let's use racial tension score (Liu 2010) as a covariate in EI's extended model through Ezi

- I use a factor analysis to combine four state-level variables: black density, social capital, political culture and diversity (Hero 2007)
- Factor analysis produces a factor score for each state, which is a measure of racial tension of each state
- I include this racial tension score in the EI extended model, and test its significance in Ezi

Note: With the racial tension scores as the covariate, the EI extended estimates for the five Deep South states now are much closer to the 45-degree diagonal. Overall, the extended model is improved, compared to the results of EI basic model.



# The benefits of EI Extended Model

- EI extended model is a method for minimizing measurement errors
- EI extended model is also a method for testing hypothesis

# The relationship between racial tension and white votes for Obama at the state level

