Protect Yourself from p-Hacking: 7 Things to Do to Avoid Committing Scientific Malpractice

Frederick Solt
frederick-solt@uiowa.edu

Yue Hu
yue-hu-1@uiowa.edu

Kevan Hudson
kevan-hudson@uiowa.edu

Jungmin Song
jungmin-song@uiowa.edu

Dong 'Erico' Yu
dong-yu@uiowa.edu

November 2, 2015

Abstract

Replication crisis
LaCour scandal
p-Hacking
malpractice not always so blatent or intentional: confirmation bias, garden of forking paths (see http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf)
Introduce Newman, Johnston, and Lown (2015*a*), perhaps noting the press attention it has received (e.g., http://www.psmag.com/health-and-behavior/five-studies-bernie-sanders-says-the-rich-are-deranged)

1 Ensure Reproducibility

Reproducibility as bare minimum for replication
script all work
packrat and checkpoint packages in R; version command in Stata
quote Newman, Johnston, and Lown (2015*a*) replication materials
Table 1 and 2 cannot be reproduced exactly
Table 3 cannot be reproduced at all: more parameters than observations

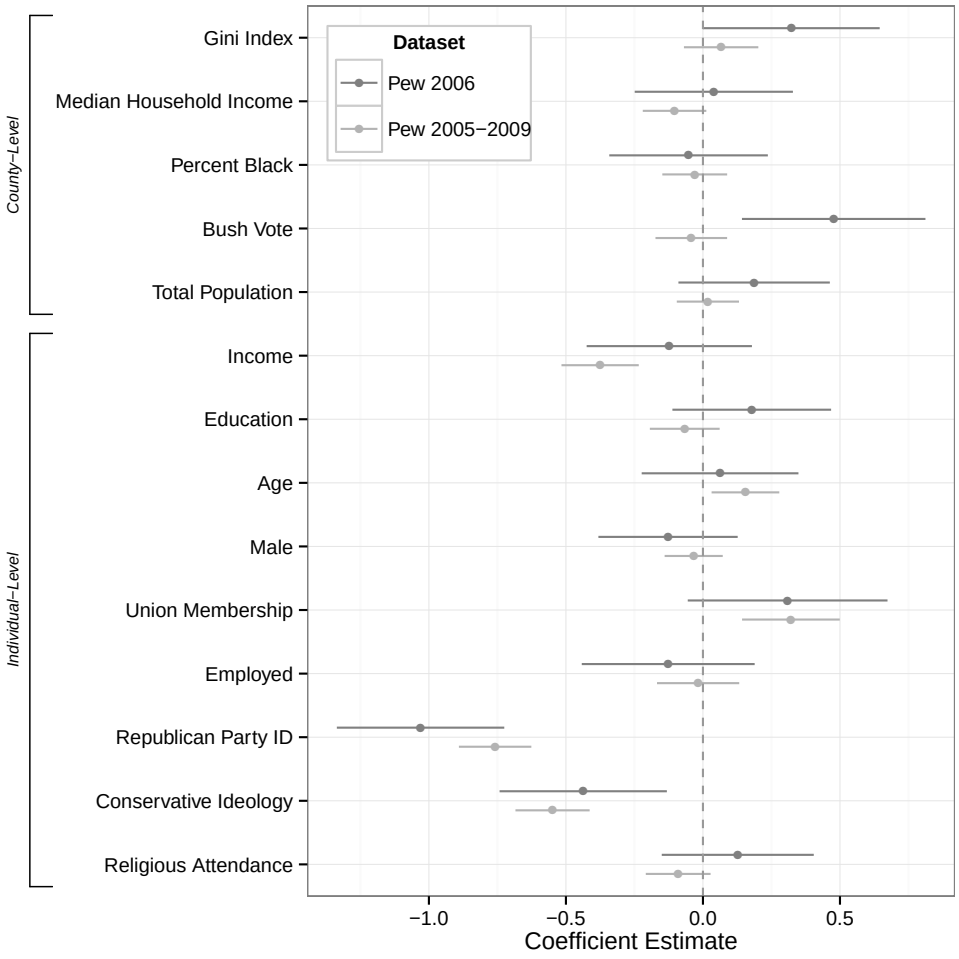
2 Work in Public

Github as baseline
discuss preregistration as option, at least for some research designs
Newman, Johnston, and Lown (2015*a*) didn't do either of these things, of course

3 Examine All Available Data

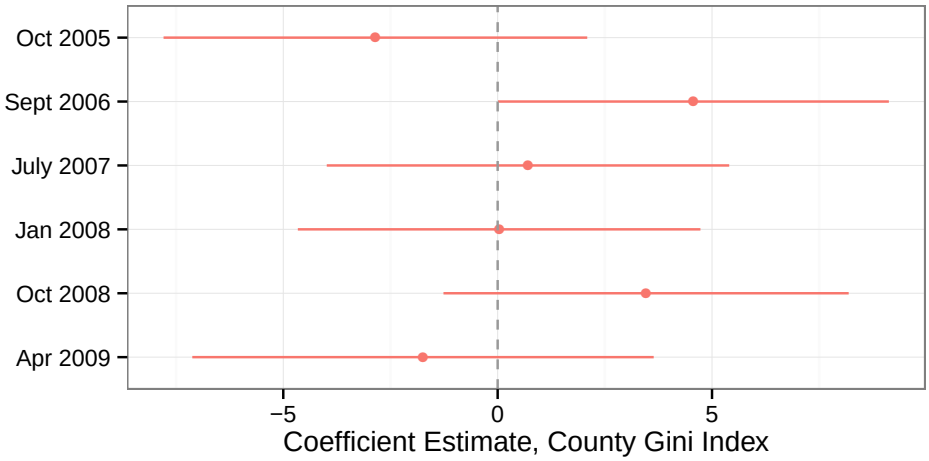
examine as much relevant evidence as possible
discuss Figure 1
discuss Figure 2

Figure 1: Local Inequality and the Perception of America as Divided into 'Haves' and 'Have-Nots': Results Using All Available Data



Notes: Results from replications of the model presented in Table 2 of Newman, Johnston, and Lown (2015*a*) on the 2006 Pew survey analyzed in that article and on pooled data from the six Pew surveys that included the same item and were conducted in the time period the article examines. The statistically significant result for county income inequality in the 2006 survey presented in that article is not evident when all of the available data are examined.

Figure 2: Local Inequality and the Perception of America as Divided into 'Haves' and 'Have-Nots': Results Using Each Available Dataset



Notes: Results for county income inequality from replications of the model presented in Table 2 of Newman, Johnston, and Lown (2015*a*) on data from each of six available surveys conducted in the in the time period examined in that article. Of the six surveys, the only one that yields a statistically significant result is the 2006 survey presented in the article.

4 Use Consistent Measures

text

5 Wrangle Data with Care

need to be really careful: double-check! Also need to be transparent.

merging: data on Bush share of vote don't match

As shown in Table 1, a quick glance at the first handful of counties analyzed in Table 1 and Table 2 (Newman, Johnston, and Lown 2015*a*) reveals that something went wrong when information on the Bush share of the vote in the 2004 election was merged into the datasets: across all counties examined in both tables, fewer than 10% have matching data, even when rounded to two decimal places. (The data analyzed in Table 1 correspond that avail-

able from other sources, so it appears that it is the Table 2 dataset that is problematic.)

Table 1: Mismatched Data on Bush Vote, from Replication Data

County	Table 1 Data	Table 2 Data
Baldwin, AL	0.76	0.79
Calhoun, AL	0.66	0.66
Chambers, AL	0.58	0.56
Cherokee, AL	0.65	0.65
Choctaw, AL	0.54	0.51
Clarke, AL	0.59	0.57

Notes: Newman, Johnston, and Lown (2015*b*) replication data on the share of the vote won by Bush in the 2004 presidential election. The first six counties, when listed alphabetically by state and county, are shown; they reveal that the data employed in Table 2 only occasionally matches that employed in Table 1, even when rounded to two digits. Overall, these data match for fewer than 10% of all counties.

coding and recoding: NJL's five point party id scale collapses leaners and weak partisans (not weak and strong partisans, and not leaners and 'true' independents). Should really use the full seven point scale; no reason to throw away that information (or to deviate from common practice)

unemployment is mismeasured in 2005, 2007, and 2009 in Table 1 due to missing employ2 variable—all of those who are not working (students, retired, etc.) are coded as unemployed

6 Multiply Impute Missing Data

Outline

missing data should be multiply imputed (e.g., King et al. 2001) (**?**)cott Long; Andrew Gelman;

church attendance—all missing are simply assigned "once or twice a month"

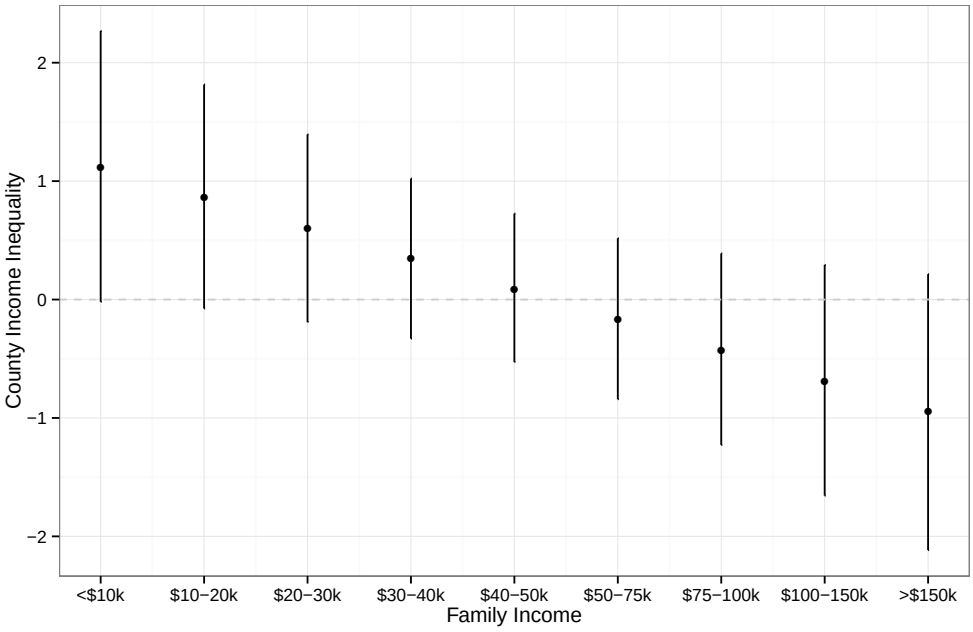
income, a variable of interest, is missing for over 10% of the sample, but values are mysteriously single-imputed (where did these values come from? they aren't meaningful–they fall between categories)

ideology, partyid also single-imputed, it seems

7 Plot Interaction Terms

It has been well known for over a decade that models containing multiplicative interaction terms require particular care in interpretation (see, e.g., Golder 2003; Braumoeller 2004; Brambor, Clark, and Golder 2006; Kam and Franzese 2007).

Figure 3: Logit Coefficients of Local Income Inequality by Respondent Income: Table 1, Model 1, From Replication Data



Notes: The coefficient for county income inequality fails to reach statistical significance for any observed level of respondent family income.

References

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58(4):807–820.

Golder, Matt. 2003. "Electoral Institutions, Unemployment, and Extreme Right Parties: A Correction." *British Journal of Political Science* 33(3):525–534.

Kam, Cindy D., and Robert J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor: University of Michigan Press.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.

Newman, Benjamin J., Christopher D. Johnston, and Patrick L. Lown. 2015*a*. "False Consciousness or Class Awareness? Local Income Inequality, Personal Economic Position, and Belief in American Meritocracy." *American Journal of Political Science* 59(2):326–340.

Newman, Benjamin J., Christopher D. Johnston, and Patrick L. Lown. 2015*b*. "Replication data for: False Consciousness or Class Awareness? Local Income Inequality, Personal Economic Position, and Belief in American Meritocracy." http://dx.doi.org/10.7910/DVN/26584, Harvard Dataverse, V2.