

Economic Inequality and Class Consciousness in the United States*

Frederick Solt Yue Hu Kevan Hudson
frederick-solt@uiowa.edu yue-hu-1@uiowa.edu kevan-hudson@uiowa.edu

Jungmin Song Dong ‘Erico’ Yu
jungmin-song@uiowa.edu dong-yu@uiowa.edu

May 5, 2016

Abstract

*Revision history and the materials needed to reproduce the analyses of this paper can be found [on Github here](#).

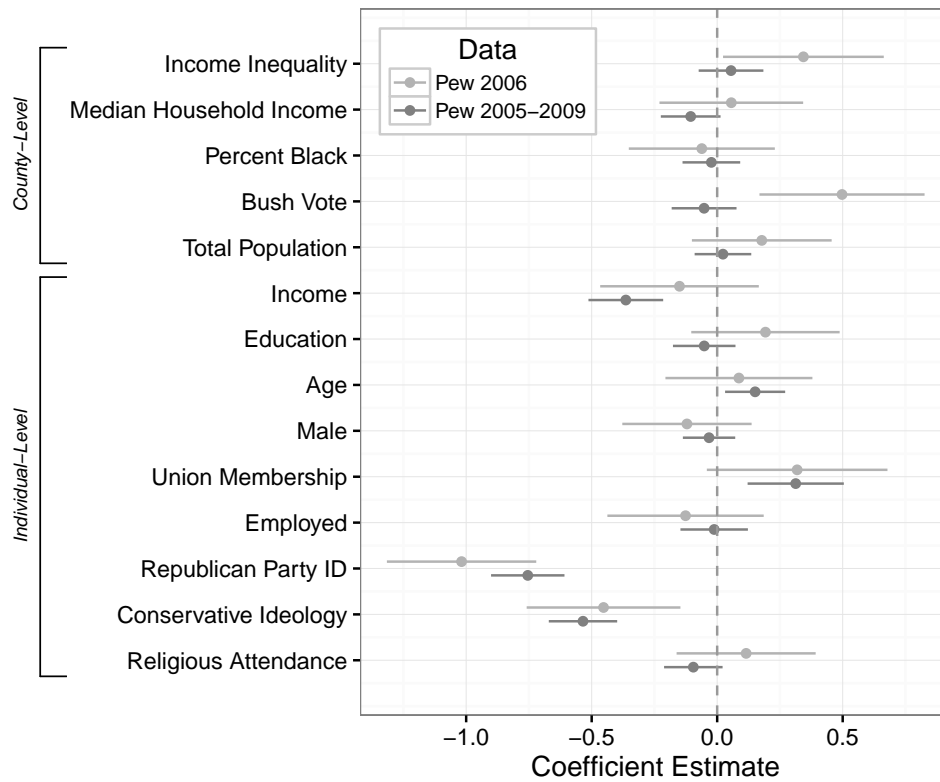
An important step in the process of conducting sound research is striving to utilize all available data. Researchers should seek to conduct their analysis on the entirety of data that they have at their disposal. Doing so provides a number of benefits: it increases the likelihood that the sample utilized captures the true distribution of the underlying population, and it affords greater leverage in testing the implications of one’s hypothesis. While the issue of selection bias cannot be avoided simply by including all available data, limiting analysis to a particular dataset, particularly when alternatives are available, may cast doubt on the inferences drawn from that analysis. By including all relevant data researchers are better able to observe the implications of their theory, thus providing greater support for the hypotheses they advance.

“an additional national data set conducted by the Pew Research Center in 2006 containing a unique set of questions tapping perceptions of economic hierarchy and inequality and respondents perception of their own position within such a hierarchy” (Newman, Johnston, and Lown 2015, 336). In reality, these questions are not unique to the 2006 dataset, but are instead present in each of the surveys the authors used in their earlier analysis. Perhaps it is by coincidence that, as shown in Figure 2, the coefficient of interest only achieves statistical significance when using the 2006 data. As illustrated, no other dataset produces a statistically significant coefficient for Gini according to the authors’ model. This provides a clear illustration of the importance of including all relevant data; failure to do so can lead to biased results.

The authors’ use of this severely truncated data has implications beyond the coefficient of interest as well. While Figure 1 clearly demonstrates that a more careful inclusion of all data produce results that run counter to the findings of Newman et al, including all

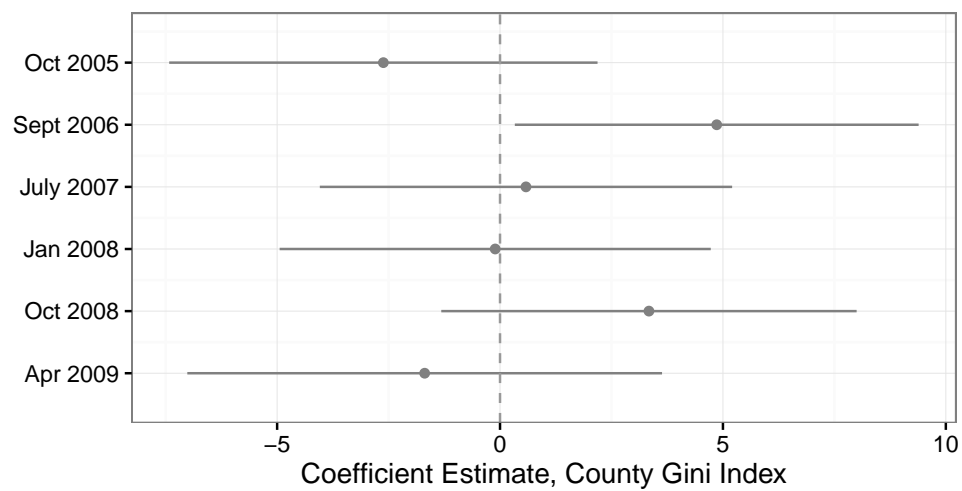
available data drastically changes the entire model, not merely the coefficient for Gini. Figure 1 provides estimates of the coefficients from Newman et al's Table 2 (p. 336) with a sample that includes data from the 2005, 2006, 2007, and 2009 surveys they use earlier in their article. When all relevant data is included, the results are drastically different. Not only does the primary variable of interest (Gini coefficient) lose statistical significance, but others do as well. Having voted for Bush is no longer a statistically significant predictor of believing America is divided into the haves and have-nots, but income becomes strongly negative and significantly associated with the same belief. Additionally, union membership gains statistical significance, indicating that belonging to a union increases the likelihood an individual perceives that have/have-not division. Ultimately, Figure 1 provides graphical representation of the dangers of not including all available data. By limiting their analysis to the sole dataset that produced a statistically significant coefficient for inequality, the authors have disguised the true relationship in order to support their theory. A properly crafted analysis reveals findings that are far less surprising; the wealthy are less likely to see America as divided into the haves and have-nots, while union members are more likely to do so. These findings counter the primary argument advanced by Newman et al, and lend strong evidence to the claim that "we should be willing to take whatever information we can acquire so long as it helps us learn about the veracity of our theory," while illustrating the pitfall of picking and choosing data that confirm our theory, while ignoring data that does not (King, Keohane, and Verba 1994, 31).

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



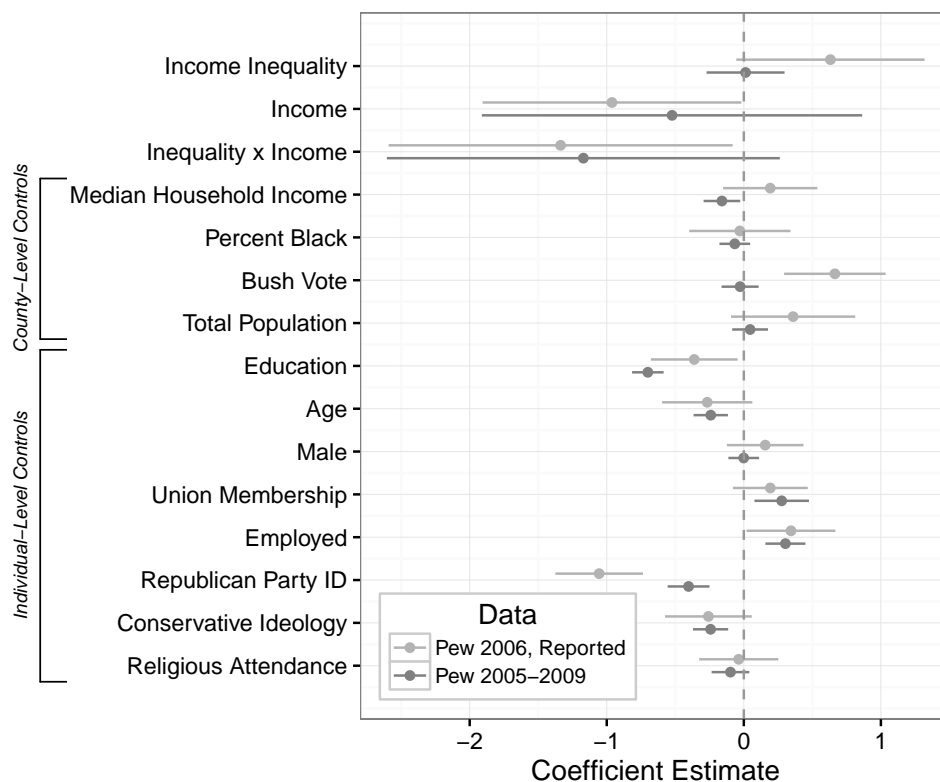
Notes: Dots indicate estimates; whiskers represent 95% confidence intervals. Results from replications of the model presented in Table 2 of Newman, Johnston, and Lown (2015) on the 2006 Pew survey analyzed in that article and on pooled data from the six Pew surveys that included the same item and were 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



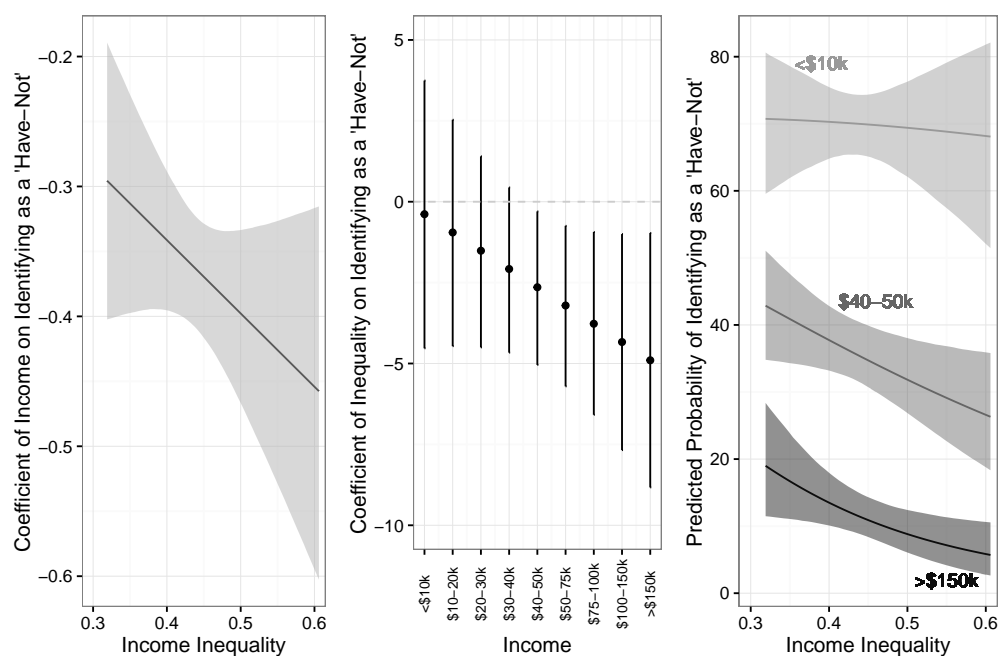
Notes: Dots indicate estimates; whiskers represent 95% confidence intervals. Results for county income inequality from replications of the model presented in Table 2 of Newman, Johnston, and Lown (2015) on data from each of six available surveys conducted in the in the time period examined in that article. Of the six surveys, the only one that yields a statistically significant result is the 2006 survey presented in the article.

Figure 3: Local Inequality and Self-Identification as a ‘Have-Not’: Results Using All Available Data



Notes: Dots indicate estimates; whiskers represent 90% confidence intervals. The model in Table 3 of Newman, Johnston, and Lown (2015), which purportedly analyzes the 2006 Pew survey, is not reproducible because it includes more parameters than observations; the reported results are depicted here. Estimating the same model on pooled data from the six Pew surveys that included the same item conducted in the time period the article yields no support for the conclusion that lower income people are more likely to identify as have-nots at any observed level of income inequality.

Figure 4: Conditional Effects of Income and Local Inequality on Self-Identification as a 'Have-Not'



Notes: Dots indicate estimates; whiskers represent 90% confidence intervals. The model in Table 3 of Newman, Johnston, and Lown (2015), which purportedly analyzes the 2006 Pew survey, is not reproducible because it includes more parameters than observations; the reported results are depicted here. Estimating the same model on pooled data from the six Pew surveys that included the same item conducted in the time period the article yields no support for the conclusion that lower income people are more likely to identify as have-nots at any observed level of income inequality.

References

- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Newman, Benjamin J., Christopher D. Johnston, and Patrick L. Lown. 2015. “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.