

It's All a Matter of Degrees: Comparing Survey and Administrative Educational Attainment Data

Andrew Foote, Larry Warren
U.S. Census Bureau

Any opinions and conclusions expressed herein are those of the author(s) and do not represent the views of the U.S. Census Bureau. All results have been approved for disclosure, DRB clearance numbers: CBDRB-FY20-CED006-0034; CBDRB-FY22-CED006-0001

Motivation

- The returns to education literature is large and extensive
- Many of these papers are based on survey data, and there have been limited efforts to measure how accurate the responses are to education questions.
- Plenty of work by done on other survey responses has found misreporting:
 - Meyer et al. (2015) shows that it is very prevalent in income assistance programs
 - Hokayem et al (2015) and Bollinger et al (2019) show how income misreporting affects earnings distributions.
- Since the ACS added the field of study question, several papers have attempted to estimate returns by field of study, with no sense of accuracy of reporting.

Contributions

- First paper to assess quality of field of study responses
- First paper to assess quality of education level responses at a large scale
- Ability to measure how these errors change over time, and how they differ by race/ethnicity and gender

Data

Degree Data

- Data are from Texas; Pennsylvania State University System; Ohio; SUNY; CUNY; Colorado
- Data report degree earned, degree field, date of graduation, institution
- Restrict to bachelor's degrees (future work is going to look at other degree levels)
- Degrees earned from 2001-2015.

American Community Survey

- Data from 2009-2016
- Focusing on the questions of highest educational attainment and field of study for bachelor's recipients.

We merge these two data sets for a 10% sample, and focus only on responses in the ACS that are taken in years after the graduation year (to ensure the respondent has received a degree)

Measuring Survey Misreporting

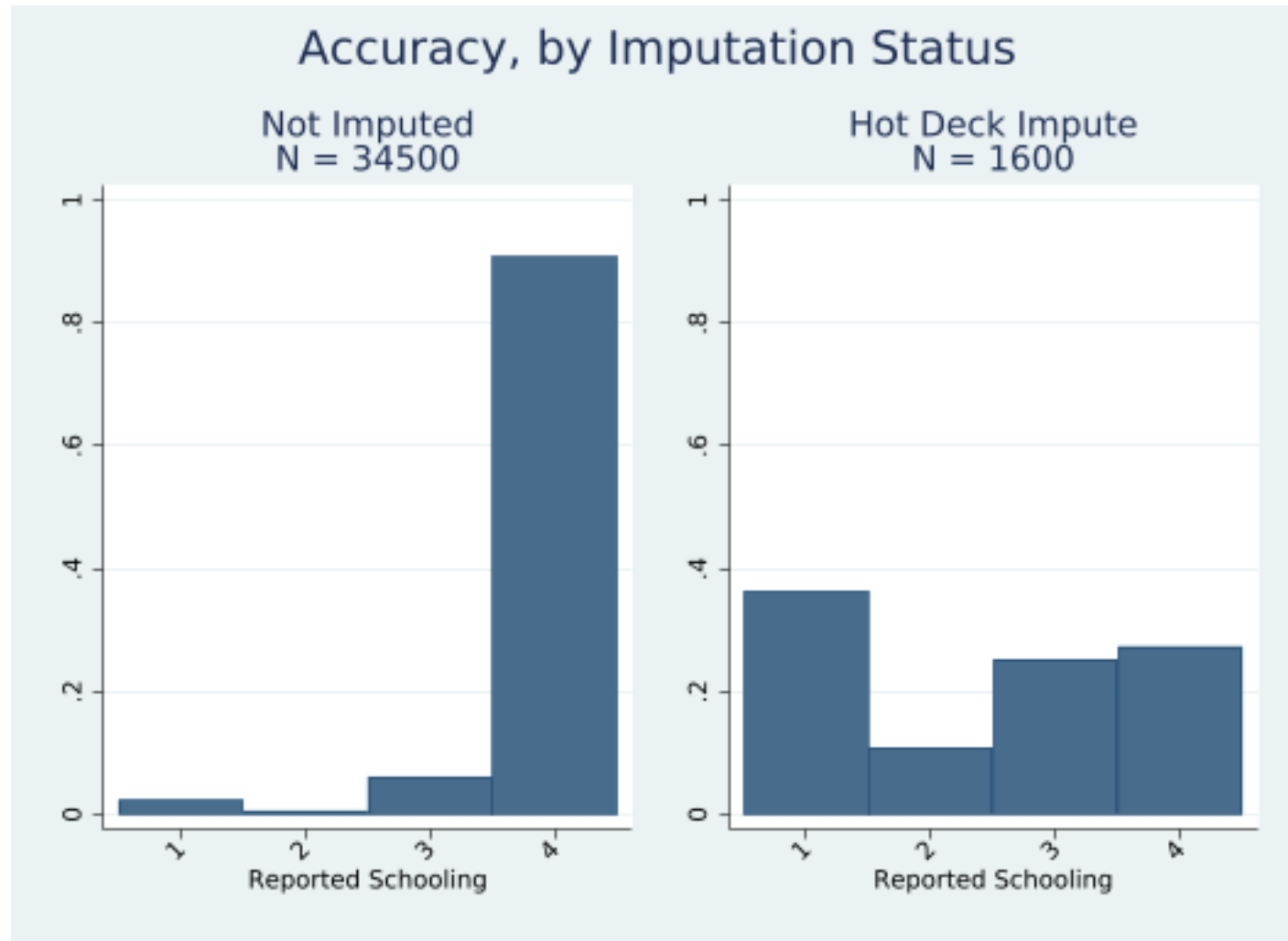
ACS Respondents are asked: What is the highest degree or level of school you have COMPLETED?

We combine these responses into four categories of responses for simplicity:

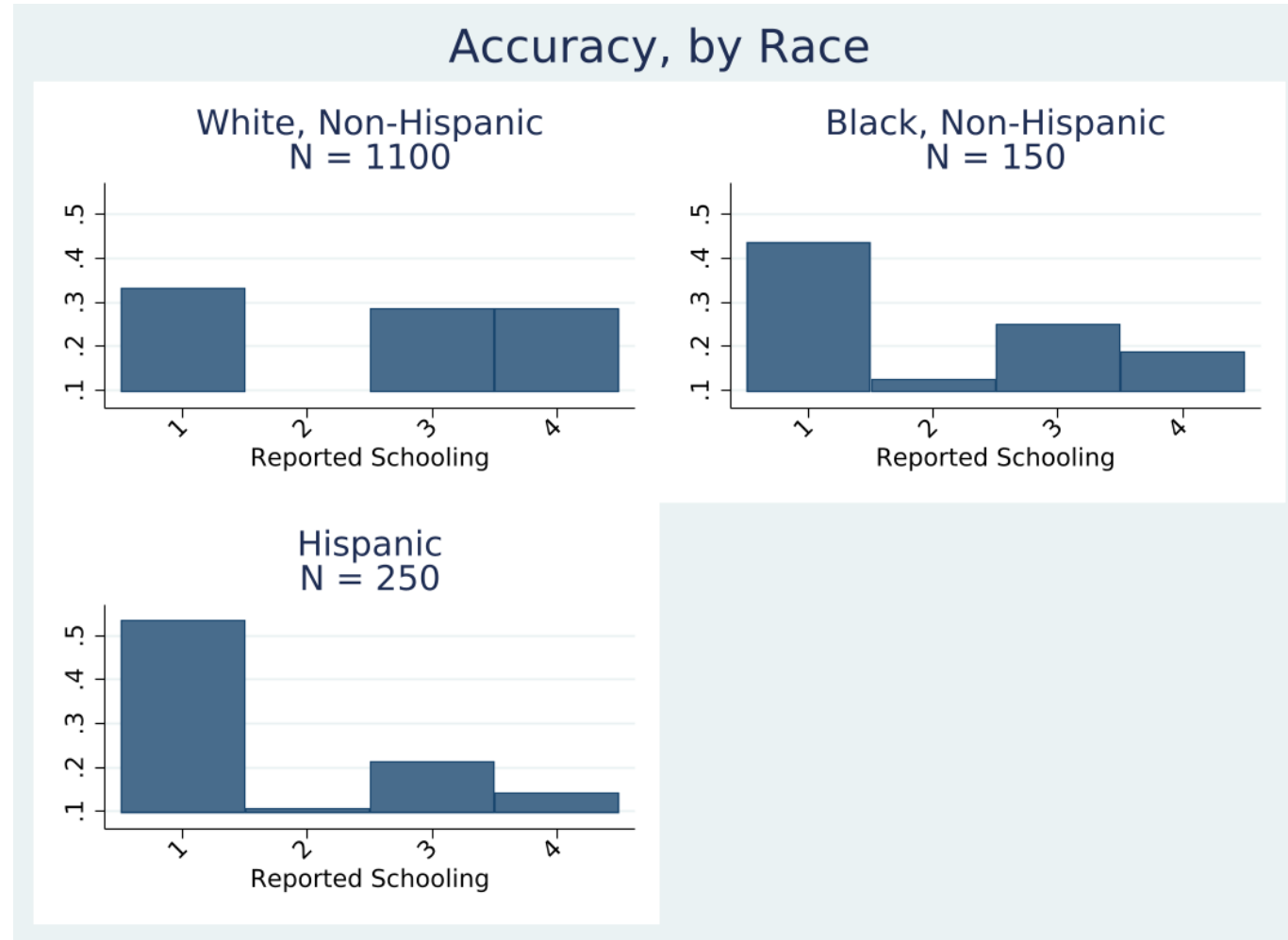
- High school dropout or below
- High school graduate
- Some college
- Bachelors or above

In our setting, since our sample only includes Bachelors recipients, the only “correct” answer is the fourth category, and we can only measure underreporting.

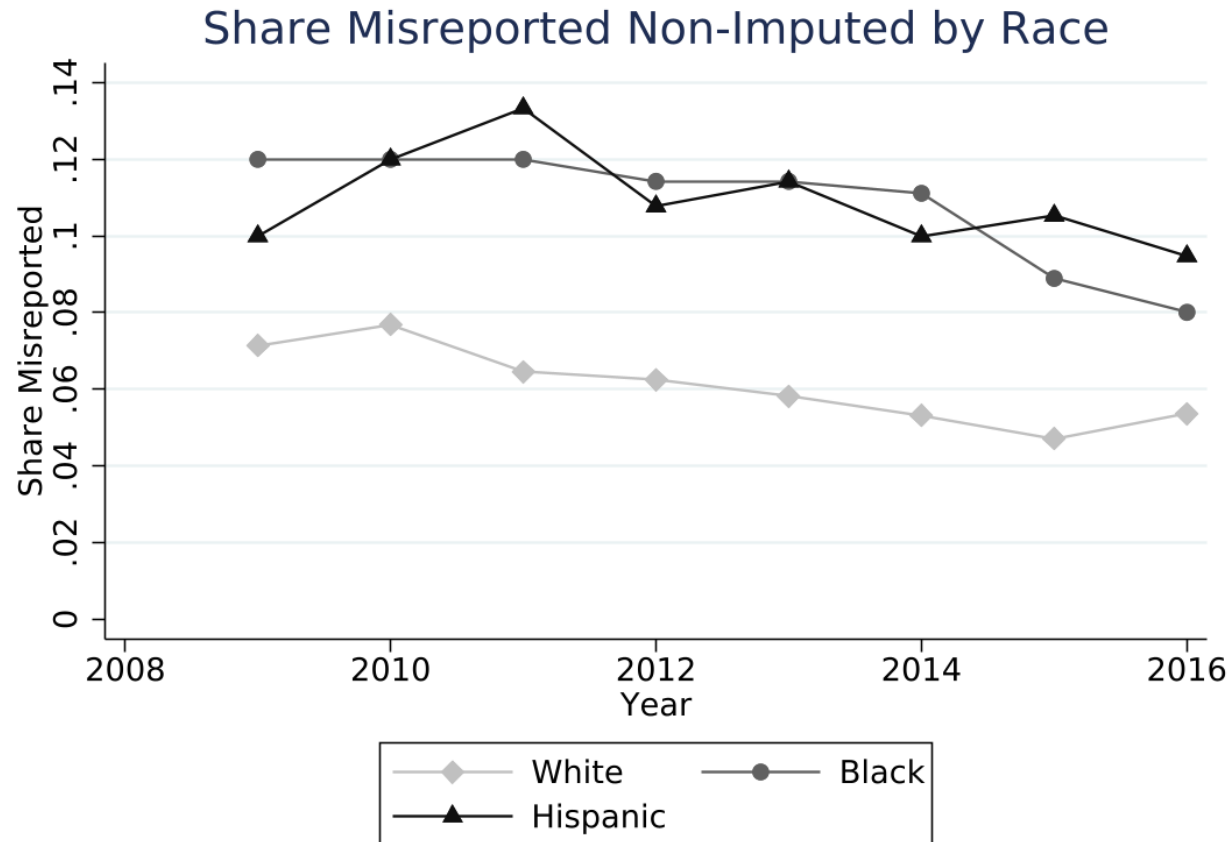
First Finding: If Response Given, High Quality



Second Finding: Differences in Impute Quality by Race

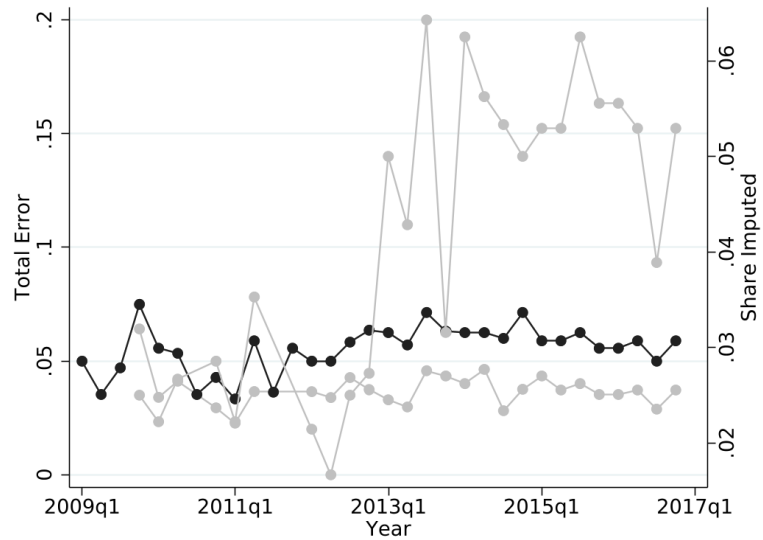


Third Finding: Conditional on Reporting, Differences in Response Quality by Race

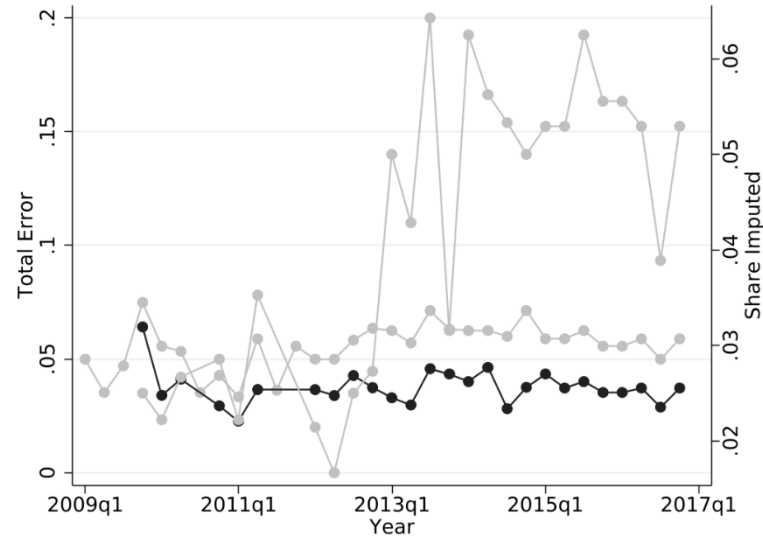


Fourth Finding: Impute Rates increased a lot in 2013

Total Error Rate

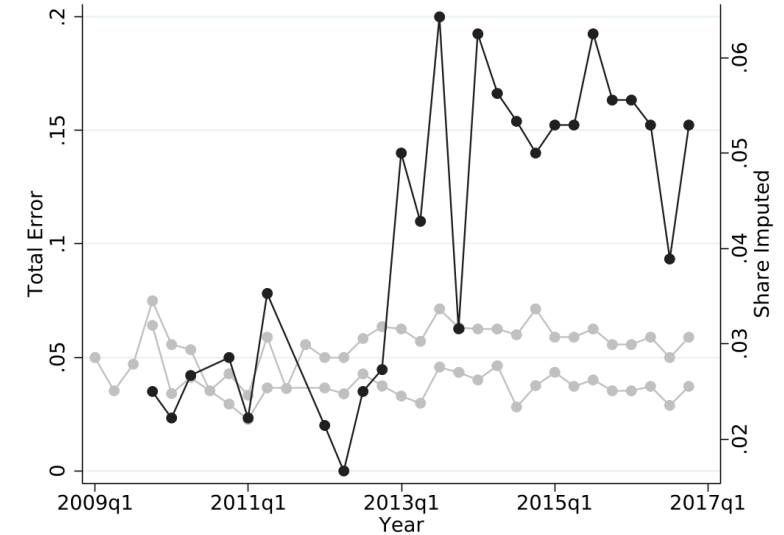


Misreported Rate



Hot Deck Rate

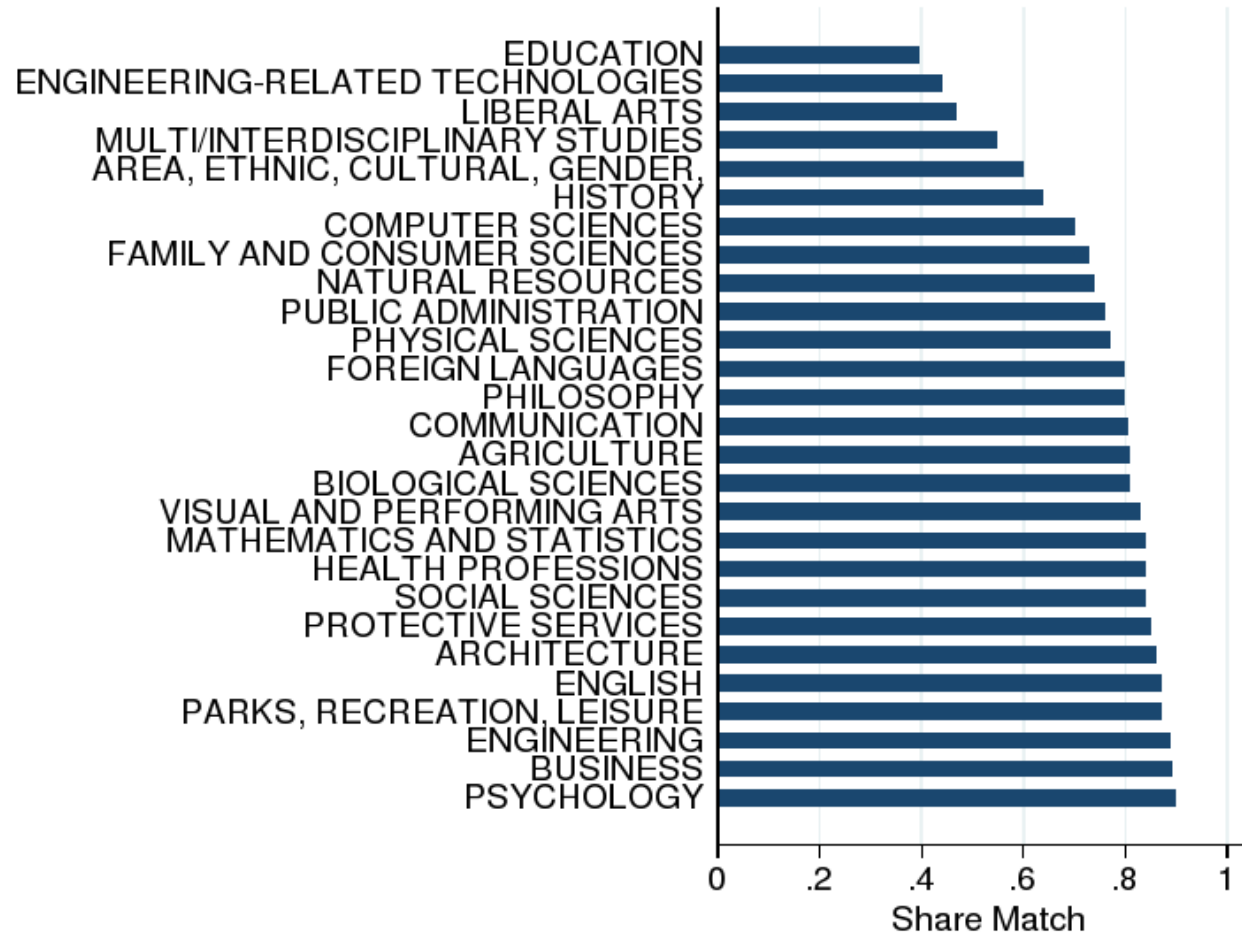
(Right Axis)



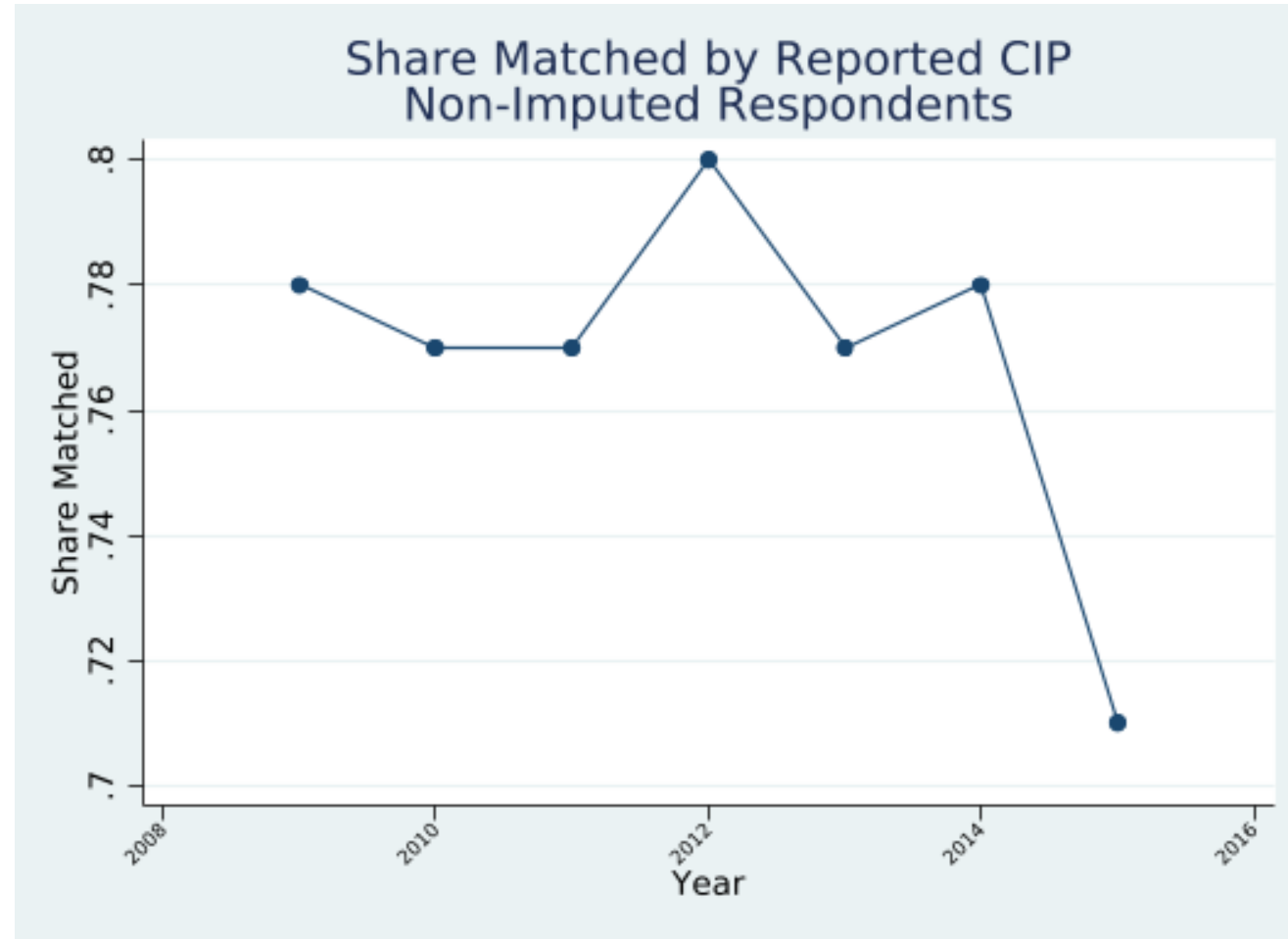
Measuring Survey Misreporting II

- We also look at the question of field of study in the ACS, and compare it administrative records.

Fifth Finding: Field of Study Mostly Reported Accurately



Field of Study Reporting Accuracy Does Not Change over Time



Conclusions

- 5-10% of ACS respondents under-report educational attainment, but about half of this error is driven by the imputation procedure.
- Misreporting and imputation rates differ significantly by race/ethnicity
- Imputation rates increase significantly in 2013, mostly due to internet and cuts to FEFU
- 20% of respondents incorrectly report field of study

Future Work

- Particularly for field of study, we want to see how misreporting may affect empirical estimates in several recent papers
- We also want to look at over-reporting, which we see for some select group of respondents
- Measure misreporting in NSCG, which should provide more accurate responses since it is more intensive.