

[Survey research in the digital age], [Probability and
non-probability sampling], [Computer-administered interviews],
[Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University



	Sampling	Interviews	Data environment
1st era	Area probability	Face-to-face	Stand-alone
2nd era	Random digital dial probability	Telephone	Stand-alone
3rd era	Non-probability	Computer-administered	Linked

Will big data kill surveys?

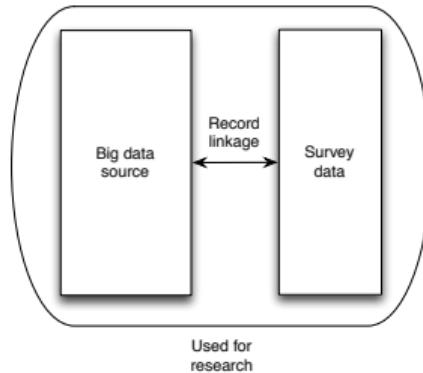


<http://schlitterblog.com/wp-content/uploads/2014/05/peanutbutterlover.jpg>

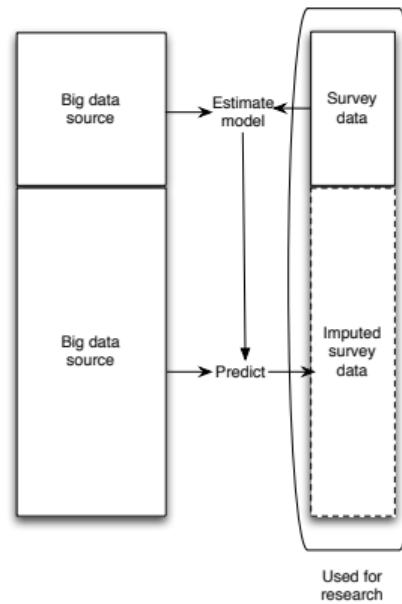


<http://schlitterblog.com/wp-content/uploads/2014/05/peanutbutterlover.jpg>

Enriched asking



Amplified asking



Note the different role of the big data in each case

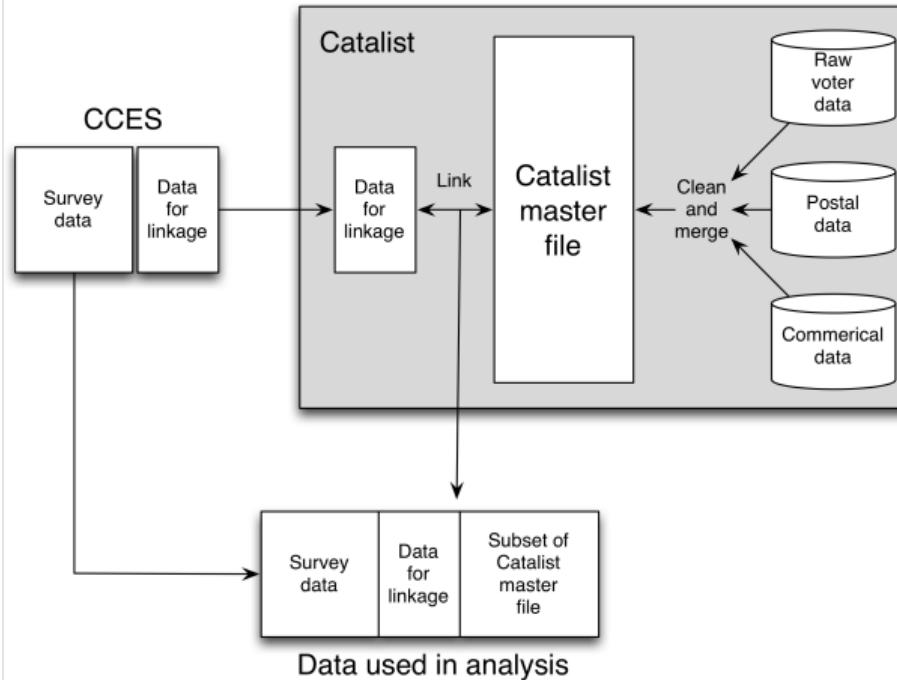
Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate

Stephen Ansolabehere
Harvard University

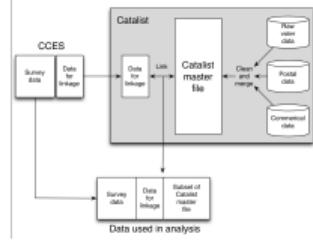
Eitan Hersh

Institution for Social and Policy Studies, Yale University, New Haven, CT 06520-8209
e-mail: eitan.hersh@yale.edu (corresponding author)

Ansolabehere and Hersh (2012)

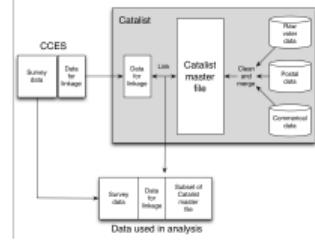


Ansolabehere and Hersh (2012)



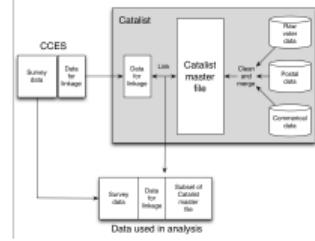
Findings:

- ▶ Over-reporting of voting is rampant: almost half of non-voters reported voting, someone who reported voting had 80% chance of actually voting



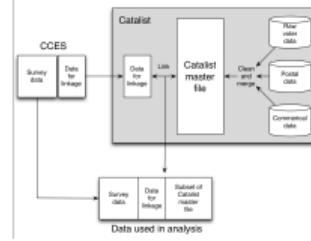
Findings:

- ▶ Over-reporting of voting is rampant: almost half of non-voters reported voting, someone who reported voting had 80% chance of actually voting
- ▶ Over-reporting is not random; it is more common among high-income, well-educated, partisans engaged in civic affairs



Findings:

- ▶ Over-reporting of voting is rampant: almost half of non-voters reported voting, someone who reported voting had 80% chance of actually voting
- ▶ Over-reporting is not random; it is more common among high-income, well-educated, partisans engaged in civic affairs
- ▶ Because of systematic over-reporting, differences between voters and nonvoters are smaller than they appear from surveys

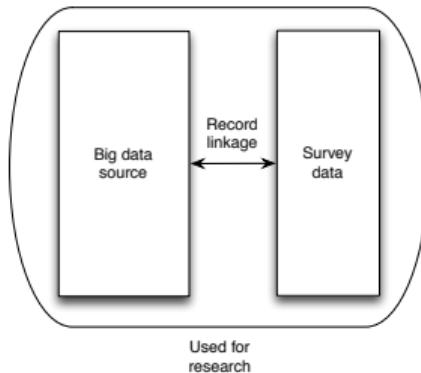


Findings:

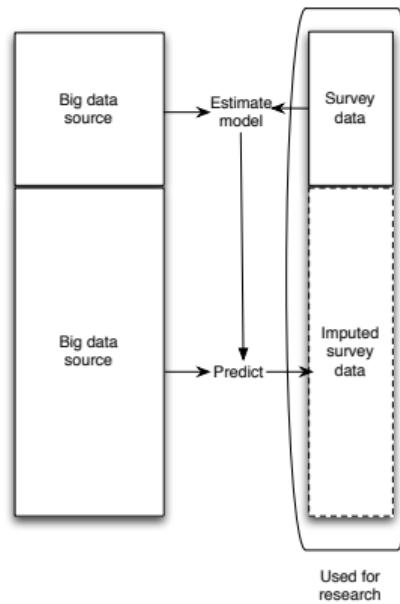
- ▶ Over-reporting of voting is rampant: almost half of non-voters reported voting, someone who reported voting had 80% chance of actually voting
- ▶ Over-reporting is not random; it is more common among high-income, well-educated, partisans engaged in civic affairs
- ▶ Because of systematic over-reporting, differences between voters and nonvoters are smaller than they appear from surveys
- ▶ Existing theories are better at predicting who will reporting voting than who will actually vote

[Anscombe and Hersh \(2012\)](#)

Enriched asking



Amplified asking



Note the different role of the big data in each case

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,^{1*} Gabriel Cadamuro,² Robert On³

Blumenstock, Cadamuro, and On (2015)

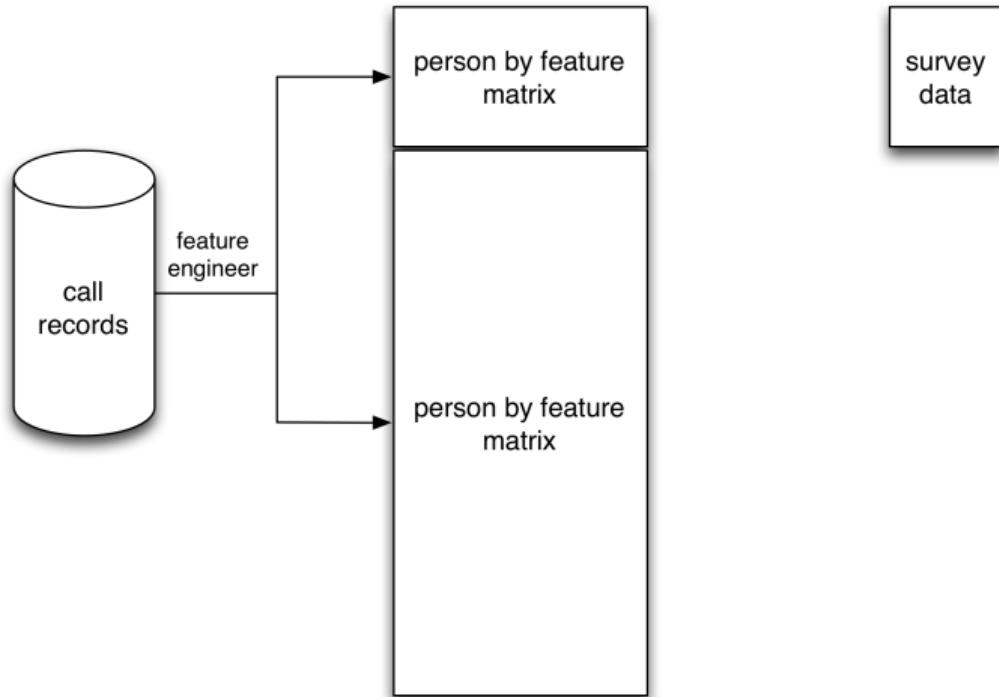


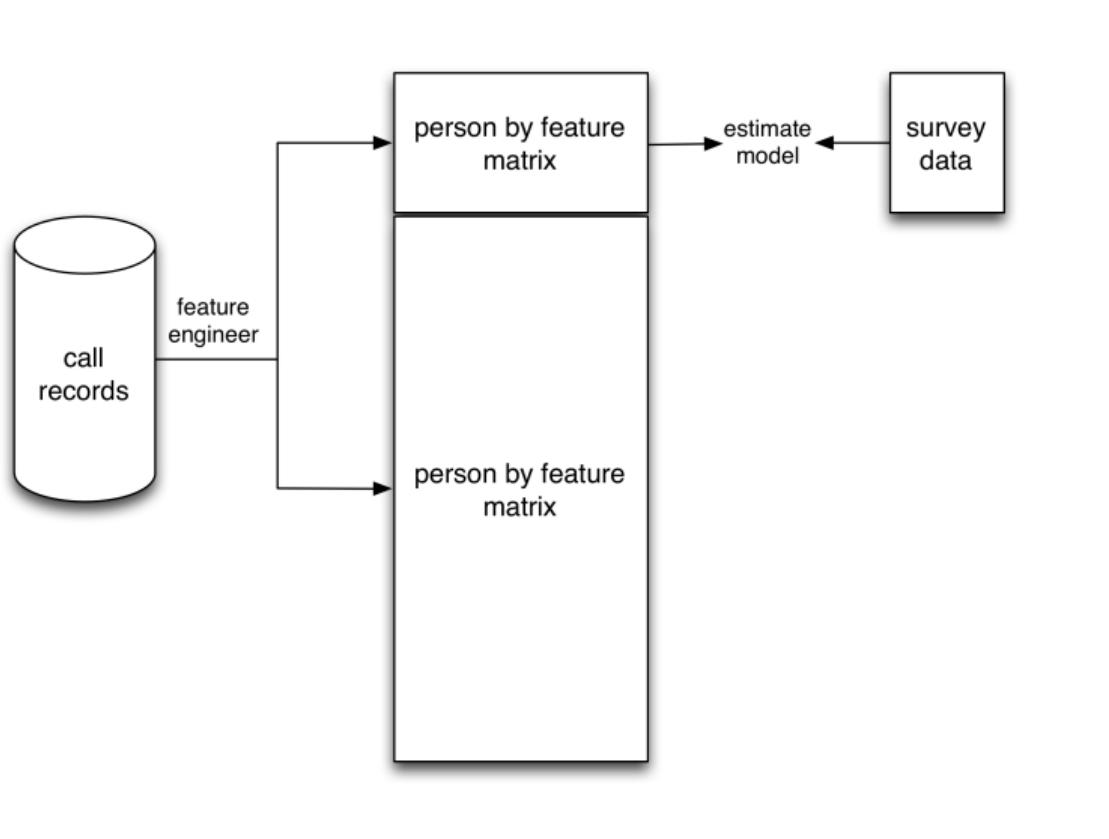


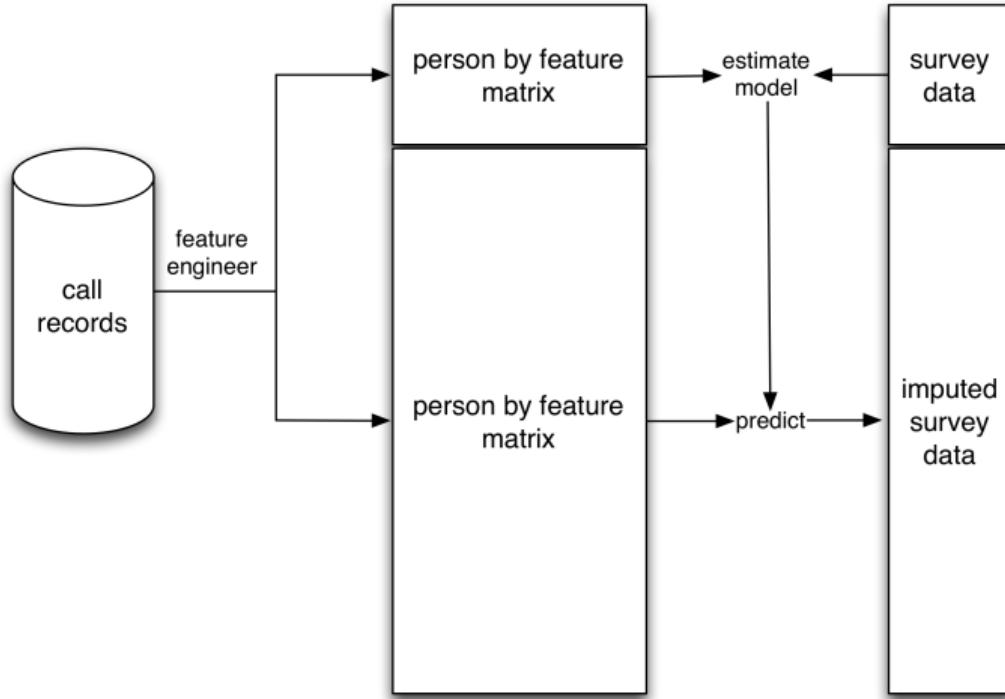
call
records

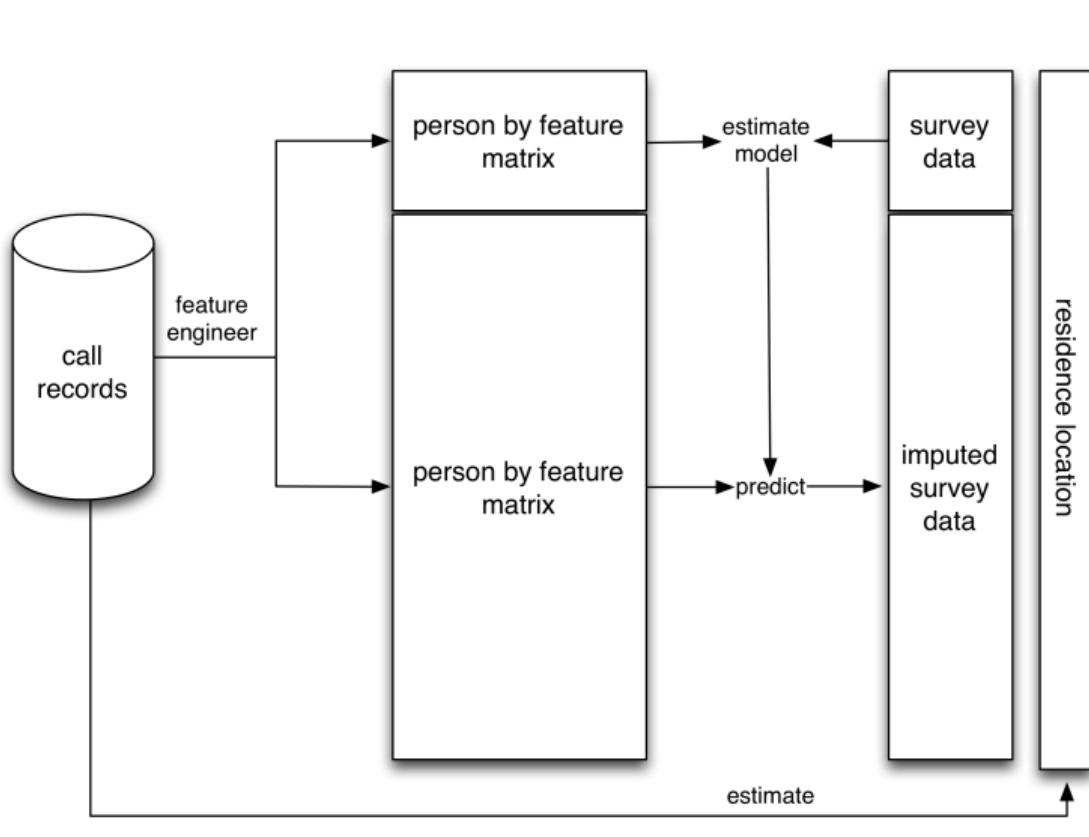


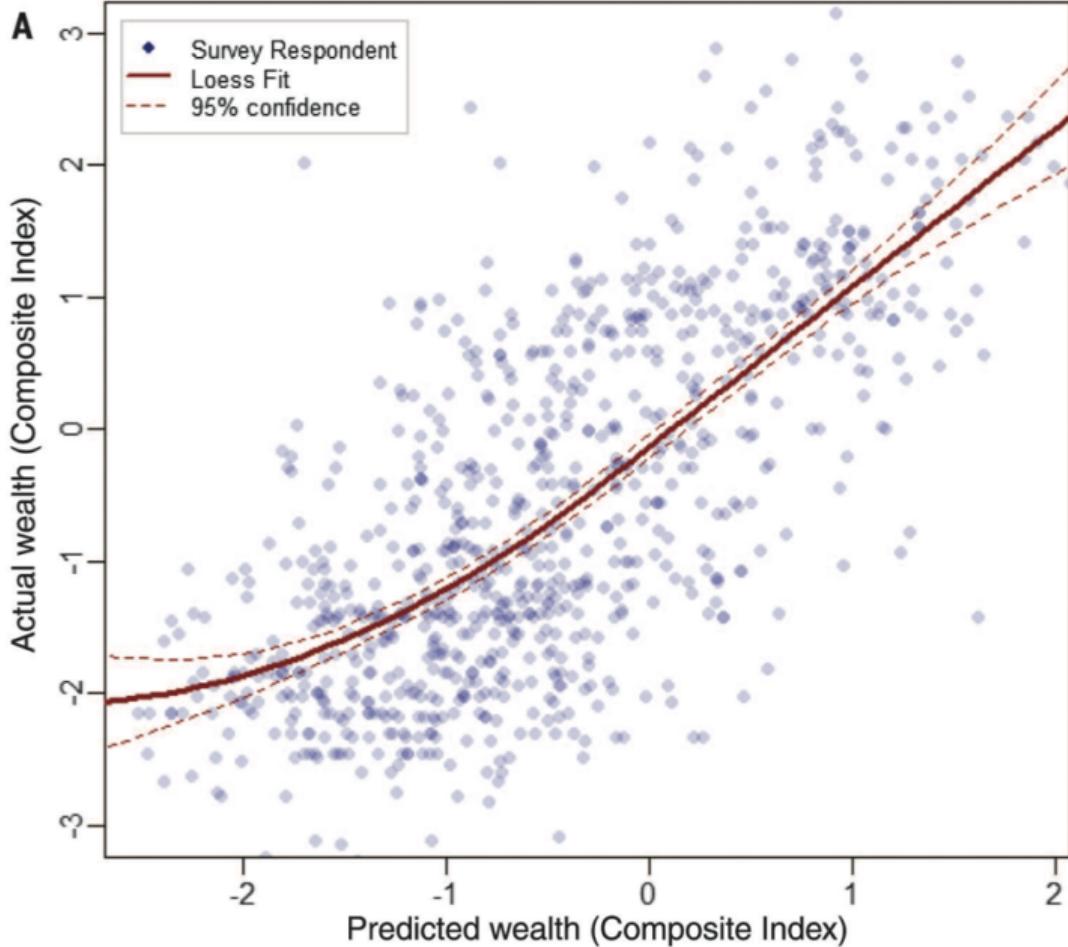
survey
data

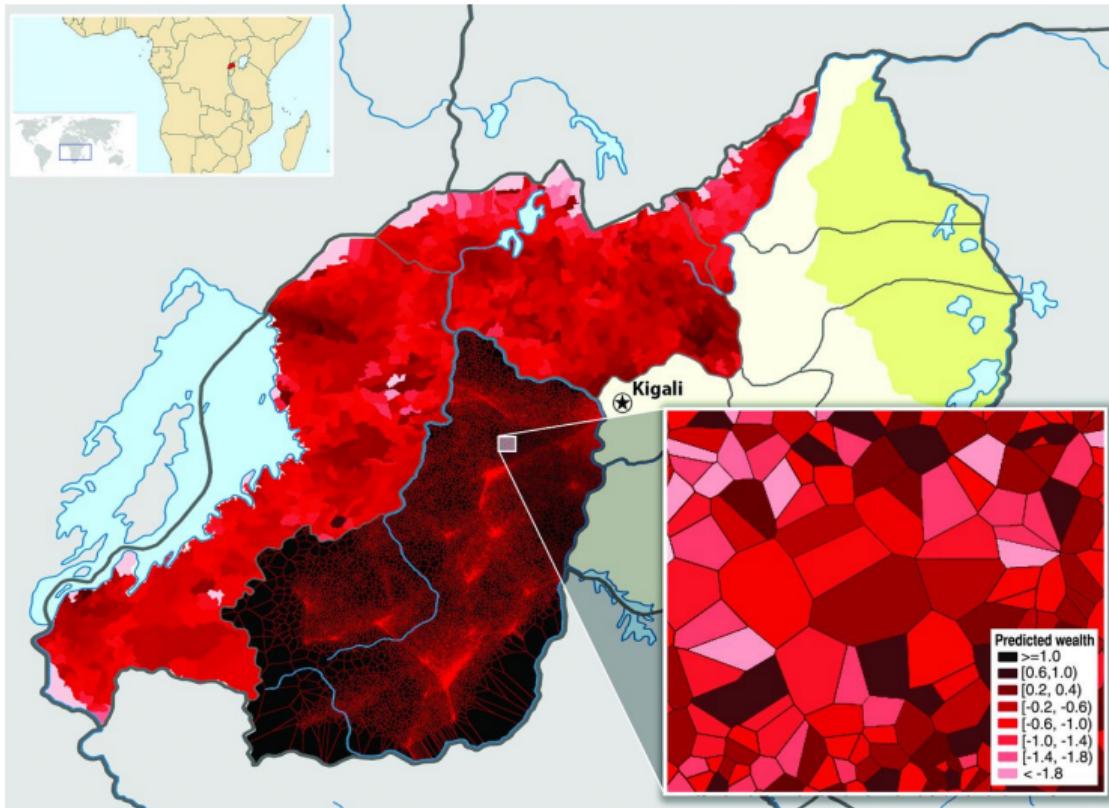


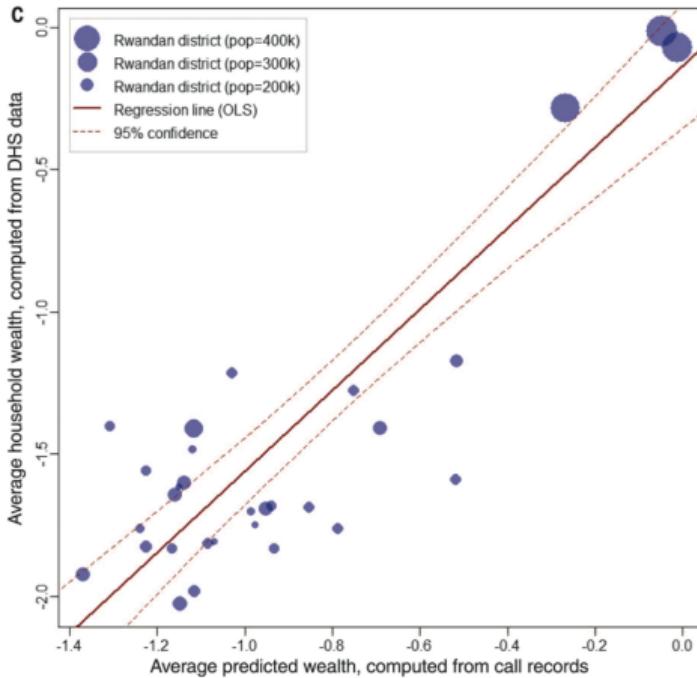


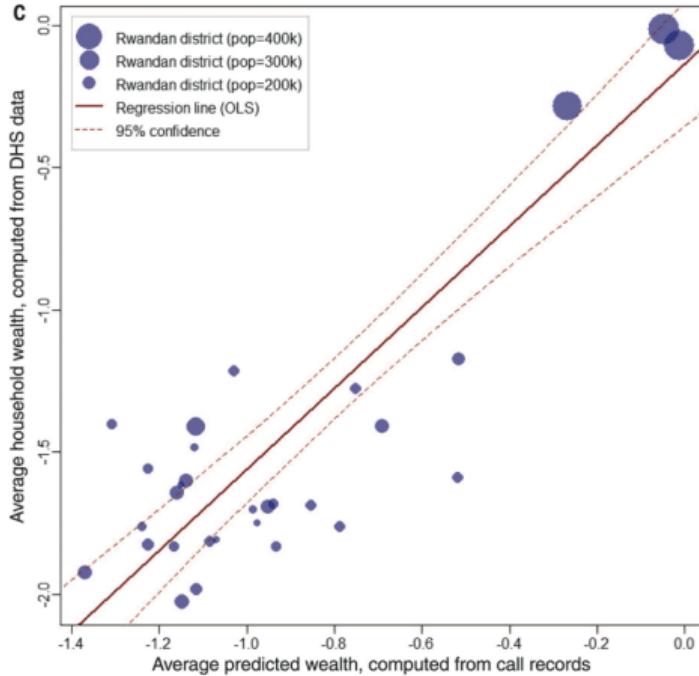












- ▶ 10 times faster
- ▶ 50 times cheaper



Readymades

+



Custommades



Wrap-up:

- ▶ Surveys and big data are compliments not substitutes

Wrap-up:

- ▶ Surveys and big data are compliments not substitutes
- ▶ Sometime we do “enriched asking” and sometimes “amplified asking” (role of big data source is different in both cases)

Wrap-up:

- ▶ Surveys and big data are compliments not substitutes
- ▶ Sometime we do “enriched asking” and sometimes “amplified asking” (role of big data source is different in both cases)
- ▶ Learn more: see “what to read next” in Ch 3 of *Bit by Bit*.

[Survey research in the digital age], [Probability and
non-probability sampling], [Computer-administered interviews],
[Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University

