A Delicate Transaction: The Data Quality Program for the 2016 and 2019 Survey of Consumer Finances

Kate Bachtell, NORC @ University of Chicago

2019 FedCASIC
April 16, 2019
Washington, D.C.
CO-AUTHORS

Catherine Haggerty, NORC @ University of Chicago
Shannon Nelson, NORC @ University of Chicago
Kevin Moore, Board of Governors of the Federal Reserve System
Richard Windle, Board of Governors of the Federal Reserve System

The analysis and conclusions set forth are those of the authors and do not indicate concurrence by colleagues, the Board of Governors of the Federal Reserve System, or NORC.
Too much missing data.

Good job!
Collecting complex financial data from U.S. households is extremely difficult

- **Every interview is an accomplishment**

Tough to give constructive feedback about these accomplishments!
SURVEY OF CONSUMER FINANCES (SCF)

Triennial survey

Sponsored by the Board of Governors of the Federal Reserve System (FRB)

Premier source of data on U.S. household finances

Dual frame sample

Public use datasets available through FRB website
“HIGH QUALITY”

Survey data

- Precise dollar amounts
- Ranges

Data Quality Feedback

- Relevant
- Informed
- Coherent
- Actionable
- Impactful
CORE FEATURES OF SCF DATA QUALITY PROGRAM

- Validation
- Monitoring
- Continuous Learning
- Feedback
- Evaluation

SCF Data Quality Program
RESULTS FROM 2016 SCF
What drives the probability of having missing dollar values?
METHODS

• Total 2016 SCF Completed Interviews (n=6,261)

• Area Probability Sample (n=4,756)

• Unweighted

• Descriptive statistics, bivariate and multivariate logistic regression
VARIABLES

- Dependent Variables:
  - Percent missing dollar values (DK/REF)
  - Any missing dollar values
VARIABLES

• Independent Variables:
  • Region of country
  • Number of refusals at week 11
  • FI assessment of Respondent:
    • Unlikely to participate (at week 11)
    • Interest (during interview)
    • Suspicion (before & after interview)
  • Period of interview completion (early/middle/late)
  • Self-rated financial knowledge
  • Financial literacy
70% have ZERO missing dollar values
27% have less than a quarter missing
Only 3% have a quarter or more missing

...So it's a small slice of the pie. But can we predict this behavior?
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region of country</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Number of refusals</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>Unlikely to participate</td>
<td>n/a</td>
<td>+ (nearly significant)</td>
</tr>
<tr>
<td>Period of completion</td>
<td>n/a</td>
<td>+ (Late)</td>
</tr>
<tr>
<td>Interest</td>
<td>3.64</td>
<td>-</td>
</tr>
<tr>
<td>Suspicion</td>
<td>2.62</td>
<td>+</td>
</tr>
<tr>
<td>Financial knowledge</td>
<td>7.22</td>
<td></td>
</tr>
<tr>
<td>Financial literacy</td>
<td>2.17</td>
<td></td>
</tr>
</tbody>
</table>
## Multivariate Logistic Regression Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.57</td>
<td><strong>0.22</strong></td>
<td>0.21</td>
</tr>
<tr>
<td>Northeast</td>
<td>0.05</td>
<td>0.09</td>
<td>1.05</td>
</tr>
<tr>
<td>South</td>
<td>-0.01</td>
<td>0.09</td>
<td>0.99</td>
</tr>
<tr>
<td>West</td>
<td>0.02</td>
<td>0.09</td>
<td>1.02</td>
</tr>
<tr>
<td># refusals</td>
<td>0.00</td>
<td>0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>Unlikely to participate</td>
<td>0.15</td>
<td>0.12</td>
<td>1.16</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.09</td>
<td>0.09</td>
<td>0.91</td>
</tr>
<tr>
<td>Late</td>
<td>0.48</td>
<td><strong>0.19</strong></td>
<td>1.62</td>
</tr>
<tr>
<td>Interest</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.98</td>
</tr>
<tr>
<td>Suspicion</td>
<td>0.19</td>
<td>*<strong>0.04</strong></td>
<td>1.21</td>
</tr>
<tr>
<td>Financial knowledge</td>
<td>0.02</td>
<td>0.02</td>
<td>1.02</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>0.06</td>
<td>0.04</td>
<td>1.07</td>
</tr>
</tbody>
</table>

n = 4,718

-2 Log Likelihood = 5688.23
CONCLUSIONS

What drives the probability of having any missing dollar values?

Late in field period

Respondent suspicion
IMPROVEMENTS FOR 2019 SCF DATA QUALITY PROGRAM

- Coding of FRB editors’ open-ended comments
- Systematic identification & communication of exemplary interviewing
  - Pre-determined continuous learning exercises mapped to specific data quality issues
  - Monitoring completion of prescribed continuous learning activities
  - Enhanced monitoring of data quality metrics over time
Get the public-use SCF data at:
https://www.federalreserve.gov/econres/scfindex.htm
https://www.federalreserve.gov/econres/scfindex.htm

Learn more about NORC’s involvement at:
http://www.norc.org/Research/Projects/Pages/survey-of-consumer-finances-SCF.aspx
Thank You!

Kate Bachtell
Bachtell-kate@norc.org