Determining Allocation Requirements for Subsampling Nonrespondents from the Annual Survey of Manufactures

Daniel R. Whitehead
Stephen J. Kaputa
Katherine (Jenny) Thompson
Disclaimer

• This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.
Problem/Solution?

• Problem
  – Decreasing response rates add both bias and variance to total estimates.
  – “Fixed budget” no longer exists.

• Solution?
  – Adaptive design strategies
Subsampling Nonrespondents

**Benefits**

- More likely to maintain quality when budget is less than anticipated.
- May reduce survey costs.
- Possibly reduce MSE.

**Drawbacks**

- Adds an extra stage to the original sample design.
- Adds a sampling error component.
Economic Census

• Conducted every five years

• Covers eighteen non-farm sectors

• Mails forms to over 4 million establishments

• Produces industry and domain estimates
Annual Survey of Manufactures (ASM) as a proxy for the Economic Census

Similar to Econ. Census
- Questionnaire
- Establishments
- Identical Editing and Imputation Procedures
- Similar Follow-up Procedures

Different from Econ. Census
- Pareto-PPS Sample of Approx. 50,000 Establishments
- Certainty and Noncertainty Units
Current Follow-up Procedure

• All nonrespondents will receive some follow-up.

• Large units receive more persistent follow-up.
  • Example: phone calls

• Small units receive limited follow-up.
  • Example: reminder letters
Research Considerations

- Cost
- Subsampling Nonrespondent Businesses
- Variance
- Nonresponse Bias
Objective

• Determine the “optimal” subsampling interval ($K$) that balances the twin goals of...
  – Reducing data collection costs
  – Minimizing the sampling variance of the estimated total.
Constraints

- Fixed budget
- Unit response rate
- Single noncertainty units
- Systematic sample
- Single subsampling rate

Cost

Subsampling Nonrespondent Businesses

Variance
Total Estimator

• We use a modified Horvitz-Thompson estimator to estimate the total value of shipments.

• Our estimator is the sum of two separate estimators.
  – The weighted sum of the initial respondents using the survey design weights
  – The reweighted sum of the subsampled units that responded to follow-up efforts
Scenario 1

ASM Sample \((n)\)

Nonrespondents \((nr_0)\)

Nonrespondent subsample

Systematic \((1/K)\) Subsample

Follow-up Round

Nonrespondents \((nr_1)\)

Nonrespondents \((nr_2)\)

\[
\vdots
\]

Nonrespondents \((nr_f)\)

Nonrespondents \((nr_{10})\)

Respondents \((r_0)\)

Respondents \((r_1)\)

Respondents \((r_2)\)

\[
\vdots
\]

Respondents \((r_f)\)

Respondents \((r_{10})\)
## ASM Follow-Up Procedure

<table>
<thead>
<tr>
<th>Contact Attempt</th>
<th>Response Propensity</th>
<th>Response From Mail ($0.90/unit)</th>
<th>Response by Phone ($5.60/unit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Mail-Out</td>
<td>35%</td>
<td>21%</td>
<td>14%</td>
</tr>
<tr>
<td>($2.75/unit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Follow-up</td>
<td>30%</td>
<td>18%</td>
<td>12%</td>
</tr>
<tr>
<td>($0.75/unit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Follow-up</td>
<td>20%</td>
<td>12%</td>
<td>8%</td>
</tr>
<tr>
<td>($2.75/unit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Follow-up</td>
<td>18%</td>
<td>11%</td>
<td>7%</td>
</tr>
<tr>
<td>($0.75/unit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th Follow-up</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>($0.75/unit)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Average Running Cost Versus Follow-Up Attempts

Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Average Response Rate Versus Follow-Up Attempts

**Scenario 1: Subsample Nonrespondents Before Attempting Follow-up**
Average Sampling Variance Versus Follow-Up Attempts

Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Scenario 1: Subsample Nonrespondents Before Attempting Follow-up
Scenario 2

ASM Sample (n)

- Respondents (r₀)
- Nonrespondents (nr₀)

Follow-up Round 1 (Completed for ALL Units)

- Respondents (r₁)
- Nonrespondents (nr₁)

Systematic (1/K) Subsample

- Respondents (r₂)
- Nonrespondents (nr₂)

- Respondents (r₃)
- Nonrespondents (nr₃)

- Respondents (r₁₀)
- Nonrespondents (nr₁₀)
Why?

• We want to increase the response rate before subsampling.
  – Units are most likely to respond to follow-up in the first round. (30% response rate)
  – The second round is a re-mail of the initial survey package.
Expectations

Delaying Subsampling  Increase Cost
Increase Response Rates
Reduce Variance
Scenario 1: Subsample Nonrespondents Before Attempting Follow-up

Scenario 2: Subsample Nonrespondents After Attempting Follow-up
Average Sampling Variance Versus Follow-Up Attempts

Scenario 1: Subsample Nonrespondents Before Attempting Follow-up

Scenario 2: Subsample Nonrespondents After Attempting Follow-up
What happened?

- Delaying Subsampling
- Decrease Follow-up Response Rate
- Increase Reweighting Factor
- Increase Variance
Limitations of Study

Controlled Study

• We assume no differentiation in quality of responses between scenarios.

• We maintain the same follow-up procedures before and after subsampling.

• Response propensities are uniformly distributed across units.

Reality

• Systematic subsampling should increase response quality as analysts target a smaller number of cases.

• Response propensity is correlated with unit size in business surveys.
Conclusion

• Simulation approach provides a tool for assessing the cost and variance trade-offs for a given design.

• Does not capture other quality benefits, such as reduced estimation bias from the more representative sample.
Areas of Future Research

• Other estimation methods

• Other allocation methods

• Nonignorable or covariate-dependent response mechanism
Acknowledgments

- Eric Fink
- Jim Liu
- Broderick Oliver
- Robert Struble
- Edward Watkins, III
• Thank you for your time and attention.

• daniel.whitehead@census.gov