An Analysis of the Mixed Collection Modes for Business Surveys at the U.S. Census Bureau

Broderick E. Oliver
Katherine (Jenny) Thompson

2013 Federal Committee on Statistical Methodology
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.
Business Surveys at the U.S. Census Bureau

- Provide “timely and accurate” economic statistics about U.S. businesses
  - Monthly, quarterly, annual, and quinquennial
  - Totals and percentage change
    - Ex. Revenue, employment, expenditures

- Employ a variety of data collection methods
  - Ex., Mail, Internet, Telephone, Fax
  - Selection made by respondent

- Auxiliary data available from certain items
Purpose of Study

The Census Bureau is trying to increase usage of internet collection over mail-out for economic programs.

• Is usage of internet collection “across the board” or confined primarily to large units?

• Does internet collection improve “quality” over other modes?
Quality and Collection Mode (Context)

• Quality = Reporting Accuracy
  • Measure of “retained” reported data
    • After analyst review and machine editing

• Challenges in Measuring Quality
  • Business population
  • Types of units
  • Data collection mode
Business Population (Fictional)

- Skewed population (value = measure of size for unit)
  - 4 units provide ≈ 47% of value (in population)

- Stratified Design
  - Certainty units (sampling weight = 1)
  - Noncertainty units (sampling weight > 1)
Types of Units

Survey Unit

Reporting Unit

Tabulation Unit

Sampling unit
- From frame
Types of Units (Equally Realistic)

At statistical period $t$

Survey Unit

Reporting Unit #1

Reporting Unit #2

Reporting Unit #3

Tabulation Unit

Survey Unit

Reporting Unit

Tabulation Unit #1

Tabulation Unit #2

Tabulation Unit #3
Data Collection Mode

• Internet
  – Automatic validation checks at submission
  – Form-design research (large companies)

• Not internet (Paper form, Fax)
  – Post-data collection validation

• Telephone follow-up can be performed with either collection mode
Internet Collection
With Business Surveys

• Paper version of form often used to gather information
  – Single contact for company
  – Different offices/persons may be responsible for different parts of form

• Consolidated data entry
Unit Response Rate (URR)

• **Unweighted** proportion of responding units
  
  – One rate per survey

  – Based on reporting unit

  – Response process metric
Quantity Response Rate (QRR)

• Weighted proportion of an estimate obtained from directly reported data

  – One rate per key item/survey (can be several)

  – Based on tabulation units

  – Estimate “quality” measure – not a process metric
    • Denominator changes by statistical period
    • Can approach 100% if reported data are retained for largest cases
Weighted Volume Response Rate (WVRR)

• **Weighted** proportion of the estimated population size obtained from respondent units
  
  – One rate per survey
  
  – Based on tabulation units
  
  – MOS is the analysis variable
  
  – Reporting unit response classification (same as the associated reporting unit classification)
  
  – Process metric and quality measure
    • if MOS and key item(s) are positively correlated
Source of Data Item (SDI)

- **Unweighted** proportion of responding units that retain reported data for an item

  - One rate per key item/survey
  - Based on tabulation unit
  - Response process metric and quality measure
Quality Measures Interaction

- Collection Mode
  - Mail
  - Fax
  - Internet
  - Other

- Size of unit
  - Certainty = large
  - Noncertainty = not large
Case Studies

Quarterly Services Survey (QSS)
- Company or EIN
- Stratified SRS-WOR design
  - New sample selected every five years
- One key variable (revenue)
  - Few reported zeros
  - Available administrative data
- Simple questionnaire

Annual Capital Expenditures Survey (ACES)
- Company
- Stratified SRS-WOR design
  - New sample selected annually
- Nine key variables (capital expenditures)
  - Legitimate reported zeros (especially small units)
  - No administrative data
- Complex questionnaire
URR
(QSS and ACES)

• Certainty Units URR > Noncertainty Units URR
  – Affects total program URR

• Units more likely to respond by internet for QSS
  (statistical period: 2009Q1 – 2011Q4)

• Units more likely to respond by form for ACES
  (statistical period: 2006 – 2010)
  – Response by internet higher in 2011
QRR by Size of Unit

**QSS** (revenue)

- Noncertainty > Certainty
- Administrative data available for large cases

**ACES** (total capital expenditures)

- Noncertainty < Certainty
- Administrative data are **not** available
  - Analysts strive to obtain reported data from large cases
QRR by Size of Unit and Collection Mode

**QSS** (revenue)

- Revenue collected via internet accounts for largest share of the QRR, regardless of size of unit

**ACES** (total capital expenditures)

- Total capital expenditures collected via internet accounts for larger share of QRR for certainty cases
# WVRR vs. URR by Data Collection Mode (ACES)

<table>
<thead>
<tr>
<th>Year</th>
<th>Form</th>
<th>Internet</th>
<th>Fax</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WVRR</td>
<td>URR</td>
<td>WVRR</td>
<td>URR</td>
</tr>
<tr>
<td>2007</td>
<td>35.4%</td>
<td>42.3%</td>
<td>41.3%</td>
<td>28.8%</td>
</tr>
<tr>
<td>2008</td>
<td>30.7%</td>
<td>38.1%</td>
<td>49.4%</td>
<td>34.4%</td>
</tr>
<tr>
<td>2009</td>
<td>28.5%</td>
<td>39.7%</td>
<td>52.2%</td>
<td>30.6%</td>
</tr>
<tr>
<td>2010</td>
<td>27.3%</td>
<td>36.7%</td>
<td>58.4%</td>
<td>33.6%</td>
</tr>
<tr>
<td>2011</td>
<td>23.0%</td>
<td>33.2%</td>
<td>61.8%</td>
<td>36.6%</td>
</tr>
</tbody>
</table>

- **WVRR < URR**: Smaller companies using paper forms more than bigger companies using internet.
- **WVRR > URR**: Bigger companies using internet more than smaller companies.
SDI by Data Collection Mode

**QSS (Revenue)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Form</th>
<th>Internet</th>
<th>Fax</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011Q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ACES (Total Capital Expenditures)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Form</th>
<th>Internet</th>
<th>Fax</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• Our analysis indicates
  – Increase usage of internet collection
  – No degradation in quality over paper
  – Less usage by small businesses

• Missed opportunities?
  – Internet forms are extensively tested for large businesses
  – How about testing usability for small businesses
Acknowledgement

Jeff Dalzell
Jim Hunt
Jim Liu
Peter Miller
Aidan Smith

Contact information:
Broderick.E.Oliver@census.gov