Response Rates and Response Patterns among New Businesses: Results from the Kauffman Firm Survey (KFS)

Zhanyun Zhao
Frank Potter
Yuhong Zheng

Third International Conference on Establishment Surveys
June 20, 2007
Montréal, Québec, Canada
Overview

- KFS Background
- Sample Design
- Non-Response Status
- Summary
Kauffman Firm Survey

- Longitudinal survey of new firms
  - Study dynamics and sustainability of new firms
- Sponsored by Ewing Marion Kauffman Foundation
  - http://www.kauffman.org/
- Dun & Bradstreet as the sampling frame
Kauffman Firm Survey (Cont.)

- Pilot study (new firms founded in 2003)
- Main study (new firms founded in 2004)
  - Baseline survey (5000 firms & weighting)
  - First follow up (4000 firms)
  - Second follow up (expect 3500 firms)
Sample Design

- Stratified random sample
- Strata
  - High tech industries (3 levels)
  - Woman-owned indicator
- Over sampling for high tech firms
High Tech Firms

- BLS definition based on D&B employment
  - Used SIC rather than NAICS
- Three categories
  - High Tech
  - Medium Tech
  - Non Tech
Woman Owned Firms

- Woman owned indicator composite
  - D&B Indicator
  - CEO’s gender / prefix

- Final indicator results
  - Tested in the pilot study
  - Added 20% using CEO’s gender
## Sample Frame

<table>
<thead>
<tr>
<th>High Tech Owned</th>
<th>June Frame</th>
<th>Deceased in Nov</th>
<th>New in Nov</th>
<th>Total All</th>
<th>Total Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>188,292</td>
<td>13,439</td>
<td>62,990</td>
<td>251,282</td>
<td>237,843</td>
</tr>
<tr>
<td>High Yes</td>
<td>361</td>
<td>21</td>
<td>166</td>
<td>527</td>
<td>506</td>
</tr>
<tr>
<td>High No</td>
<td>2,232</td>
<td>123</td>
<td>1,110</td>
<td>3,342</td>
<td>3,219</td>
</tr>
<tr>
<td>Med Yes</td>
<td>4,332</td>
<td>153</td>
<td>1,215</td>
<td>5,547</td>
<td>5,394</td>
</tr>
<tr>
<td>Med No</td>
<td>18,212</td>
<td>773</td>
<td>5,902</td>
<td>24,114</td>
<td>23,341</td>
</tr>
<tr>
<td>Non Yes</td>
<td>32,016</td>
<td>2,177</td>
<td>9,951</td>
<td>41,967</td>
<td>39,790</td>
</tr>
<tr>
<td>Non No</td>
<td>131,139</td>
<td>10,192</td>
<td>44,646</td>
<td>175,785</td>
<td>165,593</td>
</tr>
</tbody>
</table>
Sample Selection and Release

- Stratified by technology (SIC) and woman owned
- Sequential selection
  - Implicit strata: employee count, zip code
- Nov sample allocation based on
  - June and Nov sample are balanced
  - Reach Target 5,000 completes
- Released in waves
## Total Sample

<table>
<thead>
<tr>
<th>High Tech</th>
<th>Woman Owned</th>
<th>June Sample</th>
<th>June Percent</th>
<th>Nov Sample</th>
<th>Nov Percent</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>23,942</td>
<td>12.7</td>
<td>8,527</td>
<td>13.5</td>
<td>32,469</td>
</tr>
<tr>
<td>High</td>
<td>Yes</td>
<td>361</td>
<td>100</td>
<td>166</td>
<td>100</td>
<td>527</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2,232</td>
<td>100</td>
<td>1,110</td>
<td>100</td>
<td>3,342</td>
</tr>
<tr>
<td>Med</td>
<td>Yes</td>
<td>1,029</td>
<td>23.8</td>
<td>237</td>
<td>19.5</td>
<td>1,266</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>4,740</td>
<td>26.0</td>
<td>1,568</td>
<td>26.6</td>
<td>6,308</td>
</tr>
<tr>
<td>Non</td>
<td>Yes</td>
<td>2,090</td>
<td>6.5</td>
<td>670</td>
<td>6.7</td>
<td>2,760</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>13,490</td>
<td>10.3</td>
<td>4,776</td>
<td>10.7</td>
<td>18,266</td>
</tr>
</tbody>
</table>
Non-response Adjustment

- CHAID and Propensity Score
- Independently within technology group
- Location model / Response model
- Key characteristics:
  - High tech
  - Woman owned
  - Census division
  - MSA
  - Industry
  - Legal status
Response Rates

- Overall weighted rates
  - Percentage located: 90.7
  - Percentage complete/located: 54.2
  - Percentage complete: 49.1

- Response patterns on key characteristics ...
Stratification Factors

Stratification factors

- High tech
- Medium tech
- Non tech
- Woman owned
- Man owned

Located
Comp/Loc
Complete
Sampling Strata

Sampling Strata

- High tech, woman owned
- High tech, man owned
- Medium tech, woman owned
- Medium tech, man owned
- Non tech, woman owned
- Non tech, man owned

Legend:
- Located
- Comp/Loc
- Complete
Geographic Region and Metropolitan Status

Geography

Located
Comp/Loc
Complete

Midwest Northeast South West MSA Non MSA

0 10 20 30 40 50 60 70 80 90 100

MATHEMATICA Policy Research, Inc.
Item Non-response

- Focus on business financial data
  - Total assets, equity, debt, profit etc.
  - Include both main and subsequent question
  - The subsequent question could have actual value or category

- High item response rate

<table>
<thead>
<tr>
<th></th>
<th>Main Q.</th>
<th>Sub. Q.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>98.9%</td>
<td>95.6%</td>
</tr>
<tr>
<td>Median</td>
<td>99.1%</td>
<td>96.7%</td>
</tr>
</tbody>
</table>
Summary

- Stratified design to get precise understanding of newly formed firms, especially high tech

- Higher response rates are:
  - Medium tech & woman owned
  - Manufacturing
  - Midwest, non-MSA
  - Sole proprietorship

- Lower response rates are:
  - Low tech & not woman owned
  - Construction
  - Northeast, MSA
  - Partnership

- High item response rates in key questions
Thanks

- Questions and paper requests:
  zzhaoo@mathematica-mpr.com

- MPR website
  http://mathematica-mpr.com