Use of Administrative Records to Reduce Burden and Improve Quality

Amy O’Hara
CNSTAT Workshop on Respondent Burden
March 8, 2016
Using administrative records in household surveys

- Reduce burden
  - American Housing Survey (AHS)
  - American Community Survey (ACS)
  - National Survey of College Graduates (NSCG)

- Improve efficiency
- Improve quality
Research topics

- **AHS**
  - Tax amount
  - Unit size
  - Lot size
  - Year built
  - Year acquired
  - Stories
  - Bathroom
  - Bedrooms
  - Total rooms

- **NSCG**
  - Earnings
  - Degree
  - Field of study
  - Employer
  - Industry
  - Occupation
Using administrative records in household surveys

- Reduce burden
- Improve efficiency
  - best time of day
  - best mode
  - frame (National Survey of Children’s Health, National Teacher and Principal Survey (NTPS))
- tracing
- Improve quality
Using administrative records in household surveys

- Reduce burden
- Improve efficiency
- Improve quality
  - AHS, NSCG
  - Survey of Income and Program Participation
  - Consumer Expenditure Survey
Authority

- Authority to access records

§6. Information from other Federal departments and agencies; acquisition of reports from other governmental and private sources

(a) The Secretary, whenever he considers it advisable, may call upon any other department, agency, or establishment of the Federal Government, or of the government of the District of Columbia, for information pertinent to the work provided for in this title.

(b) The Secretary may acquire, by purchase or otherwise, from States, counties, cities, or other units of government, or their instrumentalities, or from private persons and agencies, such copies of records, reports, and other material as may be required for the efficient and economical conduct of the censuses and surveys provided for in this title.

(c) To the maximum extent possible and consistent with the kind, timeliness, quality and scope of the statistics required, the Secretary shall acquire and use information available from any source referred to in subsection (a) or (b) of this section instead of conducting direct inquiries.
Data sources

- Examples of data sources
  - IRS, SSA, HUD, CMS, HHS
  - SNAP, WIC, TANF
  - Property value, taxes, deeds

- Previously collected information
  - Demographic characteristics
  - Housing structure characteristics
Methods

- Modeling (SIPP)
- Substitution (AHS, NTPS)
- Hybrid approaches
## Research with the ACS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Question Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone Service</td>
<td>H8g</td>
</tr>
<tr>
<td>Year Built</td>
<td>H2</td>
</tr>
<tr>
<td>Part of Condominium</td>
<td>H16</td>
</tr>
<tr>
<td>Tenure</td>
<td>H17</td>
</tr>
<tr>
<td>Property Value</td>
<td>H19</td>
</tr>
<tr>
<td>Real Estate Taxes</td>
<td>H20</td>
</tr>
<tr>
<td>Have mortgage/mortgage amount</td>
<td>H22a and H22b</td>
</tr>
<tr>
<td>Second mortgage/HELOC and payment</td>
<td>H23a and H23b</td>
</tr>
<tr>
<td>Sale of Agricultural Products</td>
<td>H5</td>
</tr>
<tr>
<td>Social Security</td>
<td>P47d</td>
</tr>
<tr>
<td>Supplemental Security Income (SSI)</td>
<td>P47e</td>
</tr>
<tr>
<td>Wages</td>
<td>P47a</td>
</tr>
<tr>
<td>Self Employment Income</td>
<td>P47b</td>
</tr>
<tr>
<td>Interest/dividends</td>
<td>P47c</td>
</tr>
<tr>
<td>Pensions</td>
<td>P47g</td>
</tr>
<tr>
<td>Residence 1 year ago and Address</td>
<td>P15a</td>
</tr>
<tr>
<td>Number of Rooms and Bedrooms</td>
<td>H7a and H7b</td>
</tr>
<tr>
<td>Facilities</td>
<td>H8a, H8b, H8c, H8d, H8e, H8f</td>
</tr>
<tr>
<td>Fuel type</td>
<td>H13</td>
</tr>
<tr>
<td>Acreage</td>
<td>H4</td>
</tr>
</tbody>
</table>

Source: Agility in Action: A Snapshot of Enhancements to the American Community Survey, 2015
## Burden for ACS questions in study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Seconds</th>
<th>Variable</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wages</td>
<td>41</td>
<td>Real Estate Taxes</td>
<td>9</td>
</tr>
<tr>
<td>Interest/dividends</td>
<td>20</td>
<td>Supplemental Security Income</td>
<td>8</td>
</tr>
<tr>
<td>Residence 1 year ago and Address</td>
<td>18</td>
<td>Self Employment Income</td>
<td>8</td>
</tr>
<tr>
<td>Fuel type</td>
<td>14</td>
<td>Pensions</td>
<td>8</td>
</tr>
<tr>
<td>Number of Rooms and Bedrooms</td>
<td>13</td>
<td>Facilities</td>
<td>6</td>
</tr>
<tr>
<td>Year Built</td>
<td>11</td>
<td>Second mortgage/HELOC and payment</td>
<td>5</td>
</tr>
<tr>
<td>Tenure</td>
<td>11</td>
<td>Acreage</td>
<td>5</td>
</tr>
<tr>
<td>Property Value</td>
<td>11</td>
<td>Part of Condominium</td>
<td>4</td>
</tr>
<tr>
<td>Have mortgage/mortgage amount</td>
<td>11</td>
<td>Phone Service</td>
<td>1</td>
</tr>
<tr>
<td>Social Security</td>
<td>10</td>
<td>Sale of Agricultural Products</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Agility in Action: A Snapshot of Enhancements to the American Community Survey, 2015
# Difficulty/sensitivity of ACS questions in study

<table>
<thead>
<tr>
<th>Difficult or Sensitive Variables</th>
<th>Additional Variables in Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Rooms and Bedrooms</strong></td>
<td><strong>Acreage</strong></td>
</tr>
<tr>
<td><strong>Property Value</strong></td>
<td><strong>Fuel type</strong></td>
</tr>
<tr>
<td><strong>Real Estate Taxes</strong></td>
<td><strong>Have mortgage/mortgage amount</strong></td>
</tr>
<tr>
<td><strong>Year Built</strong></td>
<td><strong>Part of Condominium</strong></td>
</tr>
<tr>
<td><strong>Facilities</strong></td>
<td><strong>Phone Service</strong></td>
</tr>
<tr>
<td><strong>Interest/dividends</strong></td>
<td><strong>Residence 1 year ago and Address</strong></td>
</tr>
<tr>
<td><strong>Pensions</strong></td>
<td><strong>Sale of Agricultural Products</strong></td>
</tr>
<tr>
<td><strong>Social Security</strong></td>
<td><strong>Second mortgage/HELOC and payment</strong></td>
</tr>
<tr>
<td><strong>Supplemental Security Income</strong></td>
<td><strong>Self Employment Income</strong></td>
</tr>
<tr>
<td><strong>Wages</strong></td>
<td><strong>Tenure</strong></td>
</tr>
</tbody>
</table>

Source: Agility in Action: A Snapshot of Enhancements to the American Community Survey, 2015
Year built coverage by county
ACS 2012-Corelogic BSA match

Source: Preliminary Research for Replacing or Supplementing the Year Built Question on the American Community Survey with Administrative Records, 2015
Income

Source: Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records, 2016
Very preliminary FY2016 work

- Housing variables
  - Acreage present for 91%
  - Total Property Value present for 85%
  - Real Estate Tax present for 83%
  - Number of Bedrooms present for 61%
  - Number of Rooms present for 44%

(Source: 2014 ACS housing units matched to TPD)
Challenges

- Data quality
  - How to assess (what is “truth”)
  - Coverage (use matrix sampling where lacking)
  - Concept alignment (last year’s income vs. last 12 month income)
- Matching
- Access and use issues
Challenges

- Data quality
- Matching
  - Mismatches and missed matches
  - Person and address matching
- Access and use issues
ACS 2011 PIK rates

Source: Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records, 2016
ACS 2012-Corelogic MAF and BSA matching

Source: Preliminary Research for Replacing or Supplementing the Year Built Question on the American Community Survey with Administrative Records, 2015
Challenges

- Data quality
- Matching
- Access and use issues
  - Source volatility and reliability
  - Alignment of reference periods
Volatility

National person file from third party vendor stable, but wireless file is growing over time.

Spike in TY2007 individual income tax returns due to economic stimulus rebate

Source: 2011-2014 Targus/Neustar record counts (upper) ; TY2005-TY2009 IRS individual income tax return record counts (lower)
Most tax returns are filed between February and April...

Referring to income from the previous calendar year.
Are we trading burden reduction for lower data quality

- How are we assessing this?
- What else could we do? – study data quality
  - Example - income issues
    - Income not reported to IRS
    - Previous twelve months, annual, or quarterly
    - Do we need eight components or could we use AGI
  - Blended data vs. substitution vs. as-is
Thought experiment: Could we implement records in January 2018?

- What needs to happen if we wanted to use year built from external source?
  - Prepare data for operational uses
  - Obtain OMB approval for the questionnaire changes
  - Federal register notice before the end of 2016, given our development and print contract schedules
  - Revise ACS edit and imputation processes
  - Simulate impact on 1/5 year products
  - Ensure federal agency buy-in
Thought experiment: Impact of implementing records

- Impact on releases and products
- Impact on users
  - PUMS users →
  - HUD reports, buy-in
    (85% match rate for AHS single family homes to vendor data)
Questions?