As Time Goes By: How Period Data Influence the Estimates of Recently Arrived Immigrants in the American Community Survey

Elizabeth M. Grieco, Luke J. Larsen, and Howard Hogan
U.S. Census Bureau

ACS Data Users Conference
Alexandria, Virginia
May 11-12, 2017

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.
Immigrant’s Year of Arrival

- The ACS asks: “When did this person come to live in the United States?”
- Asked of all foreign-born individuals (naturalized U.S. citizens or noncitizens).
- Responses given in specific, single years.
- Used to estimate the size and characteristics of “recent” arrivals to the United States.
Underestimation of “Most Recent” Arrivals?

Resident U.S. foreign-born population that entered in 2000 or later

Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 1-year files, internal data (weighted).
Some Possible Explanations

- Reluctance to participation in government surveys among new, unauthorized migrants (Myers, 2004).
- Circular migrants may be confused by meaning of “come to live” (Myers, 2004; Redstone and Massey, 2004).
- Coverage error in the ACS (Van Hook and Bachmeier, 2013).
An Alternate Proposal

The authors argue that the month-to-month data collection method used in the ACS program is the primary source of the apparent recent-arrival underestimates that can be observed in the data.
Structure of Presentation

- **Background**
  *Features of ACS data and multi-year files.*

- **Methodology**
  *How relative timing of arrival and interview events in the ACS affects probabilities of inclusion and leads to underestimates of newly-arrived immigrants.*

- **Demonstration**
  *Observations of underestimation patterns evident in single- and multi-year ACS data files.*
# About the Data

<table>
<thead>
<tr>
<th>Sources</th>
<th>Populations of Interest</th>
<th>Key Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 1-yr ACS internal microdata (2008-2012)</td>
<td>• Resident U.S. population</td>
<td>• Year of entry (YOE)</td>
</tr>
<tr>
<td>• 3-yr ACS internal microdata (2006-2012)</td>
<td>• Foreign-born population</td>
<td>• Year of survey (SY)</td>
</tr>
<tr>
<td>• 5-yr ACS internal microdata (2005-2012)</td>
<td>• “Newly-arrived” foreign-born population</td>
<td>• Month of survey/interview (IMO*)</td>
</tr>
</tbody>
</table>

All estimates were derived from population-controlled person weights (PWGT).

*Available via internal data only*
ACS Data Collection

- The annual sample of housing units is divided into twelve segments (monthly panels).
- Each panel has three stages of data collection:
  - Month 1: Mail and internet modes
  - Month 2: All Month 1 nonresponse to telephone (CATI)
  - Month 3: Sample of Month 2 nonresponse to personal visits (CAPI)
- Panel month and interview month are not always the same.
Interpreting Period Estimates

- The multi-year estimates are NOT midpoints of the period or averages of the single-year estimates.
- Rather, they represent the average value of a population over the full period of the survey.
Timing of Immigrant’s Arrival Relative to Interview

Two groups of foreign-born members of the ACS population sample in a given survey year:

- *Prior arrivals* – arrived prior to survey year
- *Recent arrivals* – arrived during survey year

For recent arrivals, the eligibility of being in the ACS sample is conditional upon the month of arrival relative to the month of the interview.
Immigrant’s Probability of Inclusion ($p$) in ACS Sample

Given a housing unit selected for ACS sample:

- $p_{(Prior)}$ is generally constant (excluding mobility and mortality).
- $p_{(Recent)}$ varies with timing of interview, since arrival event can take place either before or after data collection has occurred.
- Simply put: $p_{(Prior)} \approx 1 ; \ 0 < p_{(Recent)} < 1$
- These are not sample selection probabilities; the housing units have already been selected for sample.
Scenario 1: A housing unit selected for sample in April Monthly Panel, and an immigrant moves into the housing unit at some point during the survey year.

Moves in prior to data collection.

- Probability of interviewing recently arrived immigrant in household: approximately 4/12

Moves in after data collection.

- Probability of not interviewing recently arrived immigrant in household: approximately 8/12
Windows of Opportunity

Scenario 2: A housing unit selected for sample in September Monthly Panel, and an immigrant moves into the housing unit at some point during the survey year.

<table>
<thead>
<tr>
<th>Moves in prior to data collection.</th>
<th>Moves in after data collection.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of interviewing recently arrived immigrant in household: approximately 9/12</td>
<td>Probability of not interviewing recently arrived immigrant in household: approximately 3/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JAN</th>
<th>FEB</th>
<th>MAR</th>
<th>APR</th>
<th>MAY</th>
<th>JUN</th>
<th>JUL</th>
<th>AUG</th>
<th>SEP</th>
<th>OCT</th>
<th>NOV</th>
<th>DEC</th>
</tr>
</thead>
</table>

13
Formulation (Basic)

Let $A$ be the event wherein an immigrant joins a sampled household prior to its interview. Then, $\Pr(A|m) = p_{(Recent)} \approx m/12$ for $m = 1, 2, \ldots, 12$.

It follows that:

$$
\Pr(A) = \sum_{i=1}^{12} \Pr(A|m = i) \times \Pr(m = i) \approx \frac{1}{12} \sum_{i=1}^{12} p_{(Recent)}
$$

$$
= \frac{1}{12} \sum_{i=1}^{12} \frac{i}{12} = \frac{78}{144} \approx 0.5417
$$
Formulation (Refined)

When adjusting for partial-month values:

- \( p_{(\text{Recent})} \approx \frac{m-0.5}{12} \)
- \( \Pr(A) \approx \frac{1}{12} \sum_{i=1}^{12} \frac{i-0.5}{12} = 0.5 \)

So, recent immigrants whose housing units are part of the ACS sample in a given survey year have about a 50-50 chance to be counted among the sample under the monthly data collection scheme.
Underestimate of Recent Arrival Immigrant Counts

Using the \( \Pr(A) = 0.5 \) finding, it can be shown that the estimated count of survey-year immigrant arrivals in the ACS has an expected value that is downward biased by about 50 percent.

- Plausible value for single-year ACS data files under interpretation that the estimate represents contribution of person-years toward that value.
- However, this perspective does not hold when applied to multi-year ACS data files.
Compounding Bias in ACS Multi-Year Files

- The individual survey years that make up a multi-year file contribute data to the aggregate in roughly equal measure (e.g., ACS 2010, 2011, and 2012 each contribute about the same to the 2010-2012 3-year file).

- Problem occurs when estimating counts for most recent years of entry. For example, SY 2010 does not have YOE values of 2011, but it still contributes equally toward the YOE=2011 estimate.
### Recent Arrival Estimates with ACS 3-year File

<table>
<thead>
<tr>
<th>Data year</th>
<th>Year of Arrival Within Data Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td><img src="chart" alt="2010 Year Arrivals" /></td>
</tr>
<tr>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>2011</td>
<td><img src="chart" alt="2011 Year Arrivals" /></td>
</tr>
<tr>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>2012</td>
<td><img src="chart" alt="2012 Year Arrivals" /></td>
</tr>
<tr>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
</tbody>
</table>

**2012 1-year file:**
- 2012 arrivals: $100 - 50 = 50\%$ underestimation of expected arrivals

**2010-2012 3-year file:**
- 2010 arrivals: $100 - \left(\frac{50}{3} + \frac{100}{3} + \frac{100}{3}\right) = 17\%$ underestimate of expected arrivals
- 2011 arrivals: $100 - \left(\frac{0}{3} + \frac{50}{3} + \frac{100}{3}\right) = 50\%$ underestimate of expected arrivals
- 2012 arrivals: $100 - \left(\frac{0}{3} + \frac{0}{3} + \frac{50}{3}\right) = 83\%$ underestimate of expected arrivals
### Extension to ACS 5-Year File

<table>
<thead>
<tr>
<th>Data year</th>
<th>Year of Arrival Within Data Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td><img src="image" alt="2008 arrivals" /></td>
</tr>
<tr>
<td>2009</td>
<td><img src="image" alt="2009 arrivals" /></td>
</tr>
<tr>
<td>2010</td>
<td><img src="image" alt="2010 arrivals" /></td>
</tr>
<tr>
<td>2011</td>
<td><img src="image" alt="2011 arrivals" /></td>
</tr>
<tr>
<td>2012</td>
<td><img src="image" alt="2012 arrivals" /></td>
</tr>
</tbody>
</table>

**2008-2012 5-year file:**

- **2008 arrivals:** \( 100 - \left( \frac{50}{5} + \frac{100}{5} + \frac{100}{5} + \frac{100}{5} + \frac{100}{5} \right) \) = 10% underestimation
- **2009 arrivals:** \( 100 - \left( \frac{0}{5} + \frac{50}{5} + \frac{100}{5} + \frac{100}{5} + \frac{100}{5} \right) \) = 30% underestimation
- **2010 arrivals:** \( 100 - \left( \frac{0}{5} + \frac{0}{5} + \frac{50}{5} + \frac{100}{5} + \frac{100}{5} \right) \) = 50% underestimation
- **2011 arrivals:** \( 100 - \left( \frac{0}{5} + \frac{0}{5} + \frac{0}{5} + \frac{50}{5} + \frac{100}{5} \right) \) = 70% underestimation
- **2012 arrivals:** \( 100 - \left( \frac{0}{5} + \frac{0}{5} + \frac{0}{5} + \frac{0}{5} + \frac{50}{5} \right) \) = 90% underestimation
Recent Arrivals by Month of Interview (ACS 2012)

Source: U.S. Census Bureau, American Community Survey 2012 1-year file, internal data (weighted).
Year of Entry Among Single- and Multi-Year Files (ACS 2008-2012)


Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 1-year files, internal data (weighted).
Year of Entry Among 3-Year Files (ACS 2006-2012)


Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 3-year files, internal data (weighted).
Year of Entry Among 5-Year Files (ACS 2005-2012)

Summary

- Year-long ACS data collection methodology (i.e., monthly panels) results in underestimation of newly arrived immigrants.
- The method of creating multi-year datasets compounds the bias for recent-arrival estimates.
- Researchers are urged to use caution when interpreting year-of-entry data, particularly for immigrants that arrived within the last five years.
Questions?

elizabeth.m.grieco@census.gov

luke.j.larsen@census.gov

howard.r.hogan@census.gov