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

THE PROBLEM

The "year of entry" question included on the American Community Survey (ACS) asks those born abroad: "When did this person come to live in the United States?" Year of entry data often are used to estimate the size and characteristics of recent arrivals, defined as those who came to live in the United States five to ten years before the survey.
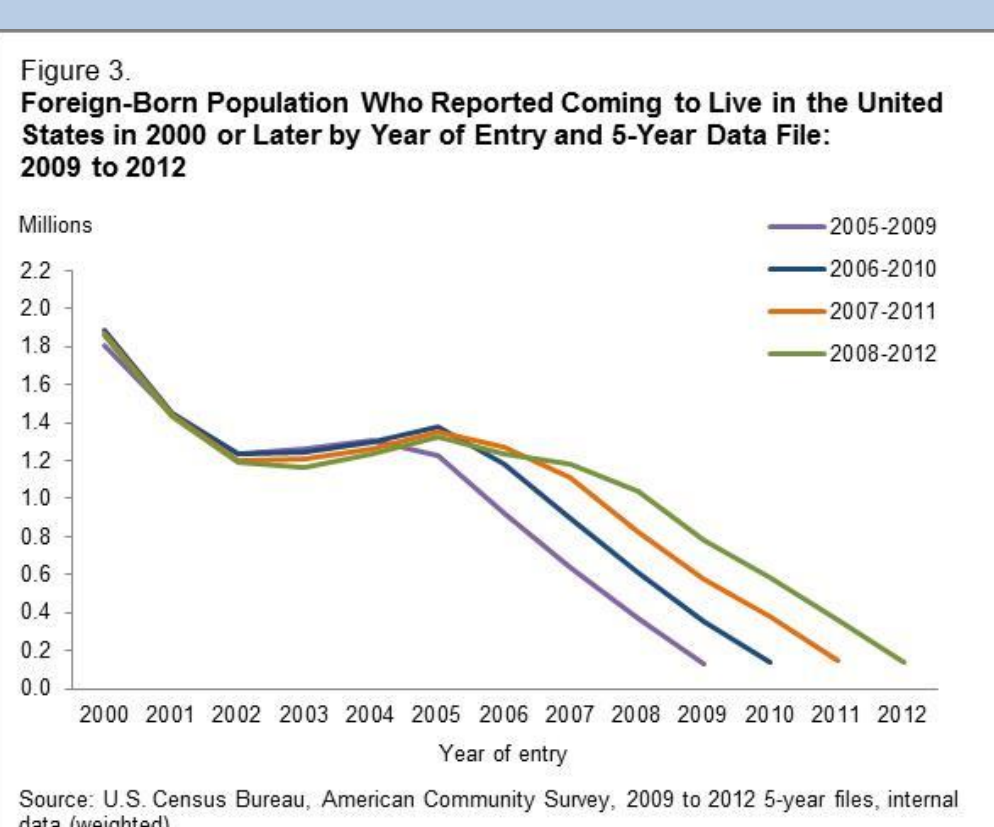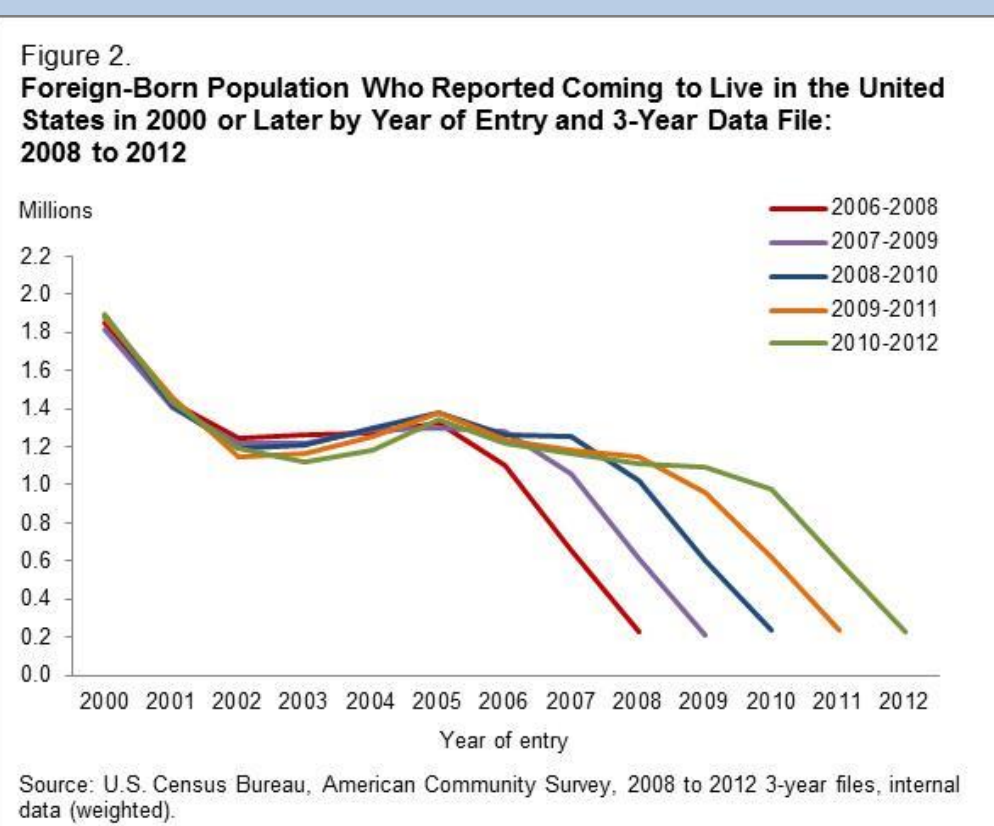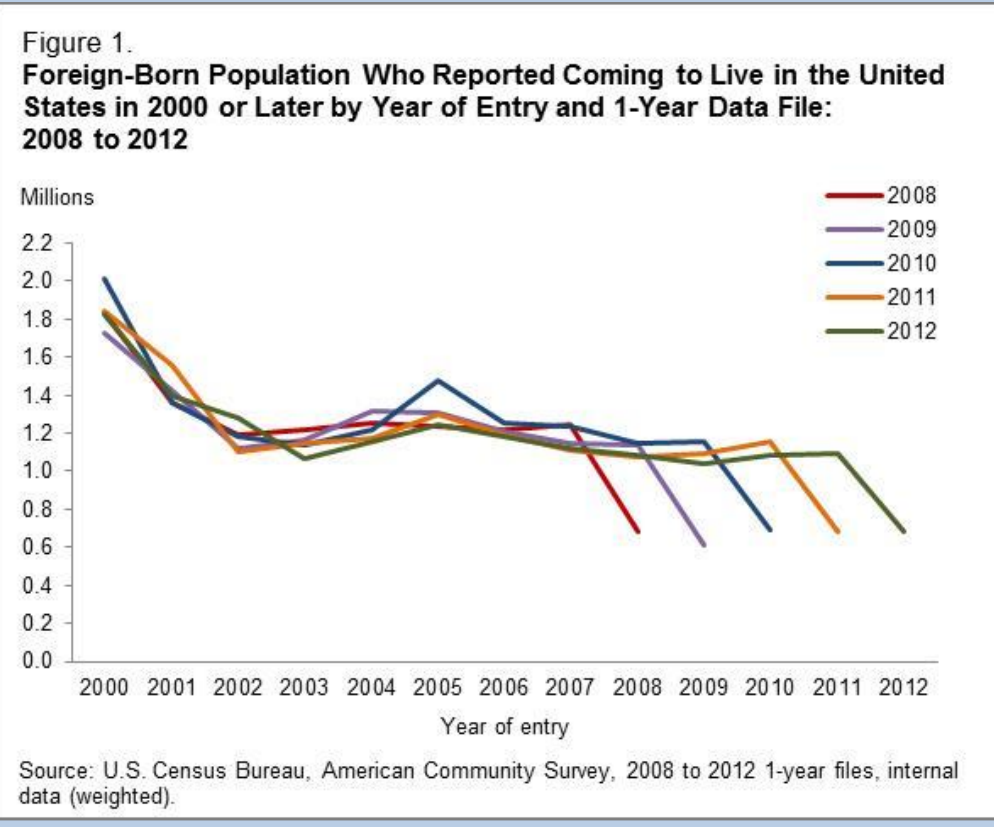
Since the inception of the ACS, the estimated count of those arriving during the survey year has been consistently less than counts from earlier years of entry, typically by about half.

Figure 1 shows that, for each 2008 to 2012 1-year file, the estimate of the final year of entry is notably lower than the previous arrival year. When viewed sequentially, however, the year of entry is "corrected" in the following year. For example, note how the 2011 year-of-entry estimate in the 1-year 2011 file increases dramatically in the 1-year 2012 file, falling in line with the general trend. This pattern also appears in the multi-year data, as shown in Figures 2 and 3, although the decline begins in the first year of the multi-year period.

Researchers have attributed low estimates of recent arrivals to various causes, such as the reluctance to participate in government surveys among new or unauthorized migrants; confusion over the meaning of "come to live" by circular migrants; and coverage error in the ACS.

OUR HYPOTHESIS

The month-to-month data collection method used in the ACS program is the primary source of the apparent recent-arrival underestimates observed in the data.

Figure 1.
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 1-Year Data File: 2008 to 2012

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

Figure 2.
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 3-Year Data File: 2008 to 2012

Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 3-year files, internal data (weighted).

Figure 3.
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 5-Year Data File: 2009 to 2012

Source: U.S. Census Bureau, American Community Survey, 2009 to 2012 5-year files, internal data (weighted).

DATA

The data are from the 1-year, 3-year, and 5-year ACS internal microdata files. The populations of interest include the resident U.S., foreign-born, and "newly-arrived" foreign-born populations. The key variables include year of entry, year of survey, and month of survey or interview. All estimates were derived from population-controlled person weights.

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 interviews
   Month 3: Sample of Month 2 nonresponse to personal visits

Note that the panel month (e.g., September) and interview month (e.g., November) are not always the same.

TIMING OF ARRIVAL RELATIVE TO INTERVIEW

In the ACS population sample in a given survey year, there are two groups of foreign born: 1) *Prior arrivals* who arrived prior to the survey year and 2) *Recent arrivals* who arrived during the 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 interview.

Figure 4.
The Influence of Period Data on the Probability that a Recently Arrived Immigrant Will Be Interviewed in the Survey Year of the American Community Survey

Housing Unit 1 in April Monthly Panel
Household receives new foreign-born arrival; data collected in April of the survey year.

Moves in prior to data collection.	Moves in after data collection.
Probability of interviewing recently arrived immigrant in household: approximately 4/12	Probability of not interviewing recently arrived immigrant in household: approximately 8/12

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |

Housing Unit 2 in September Monthly Panel
Household received new foreign-born arrival; data collected in September of the survey year.

Moves in prior to data collection.	Moves in after data collection.
Probability of interviewing recently arrived immigrant in household: approximately 9/12	Probability of not interviewing recently arrived immigrant in household: approximately 3/12

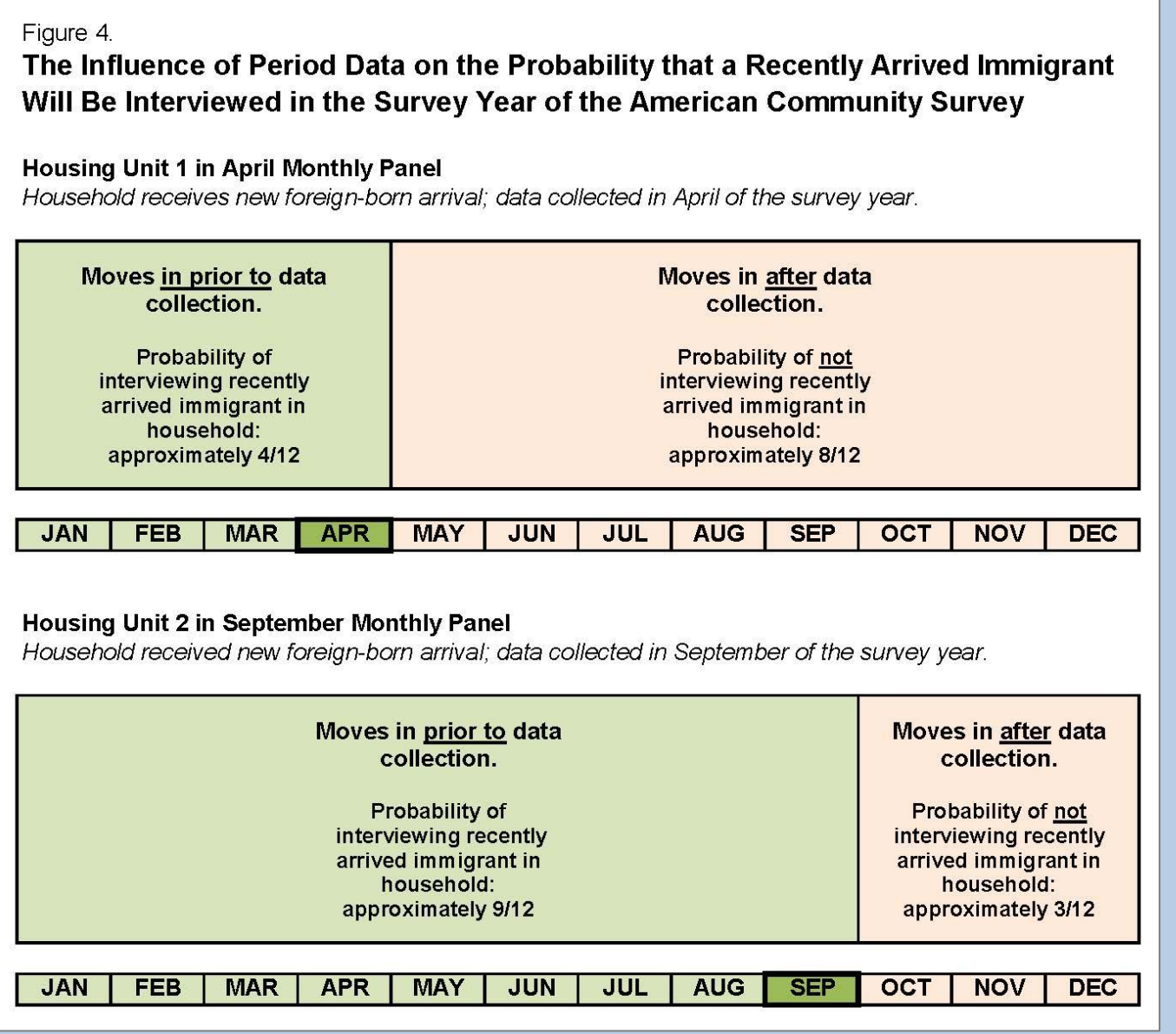| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |

Figure 4 illustrates the influence continuous data collection has on the probability that a recently-arrived immigrant will be included in sample. Our concern is when recent immigrants arrive after their sampled housing units were already interviewed and therefore are not themselves included in the data collection.

BIASED ESTIMATES OF RECENT ARRIVALS

Given a housing unit selected for the ACS sample:

$p_{(Prior)}$ is generally constant (excluding mobility and mortality) while $p_{(Recent)}$ varies with timing of interview, since the arrival event can take place either before or after data collection has occurred.

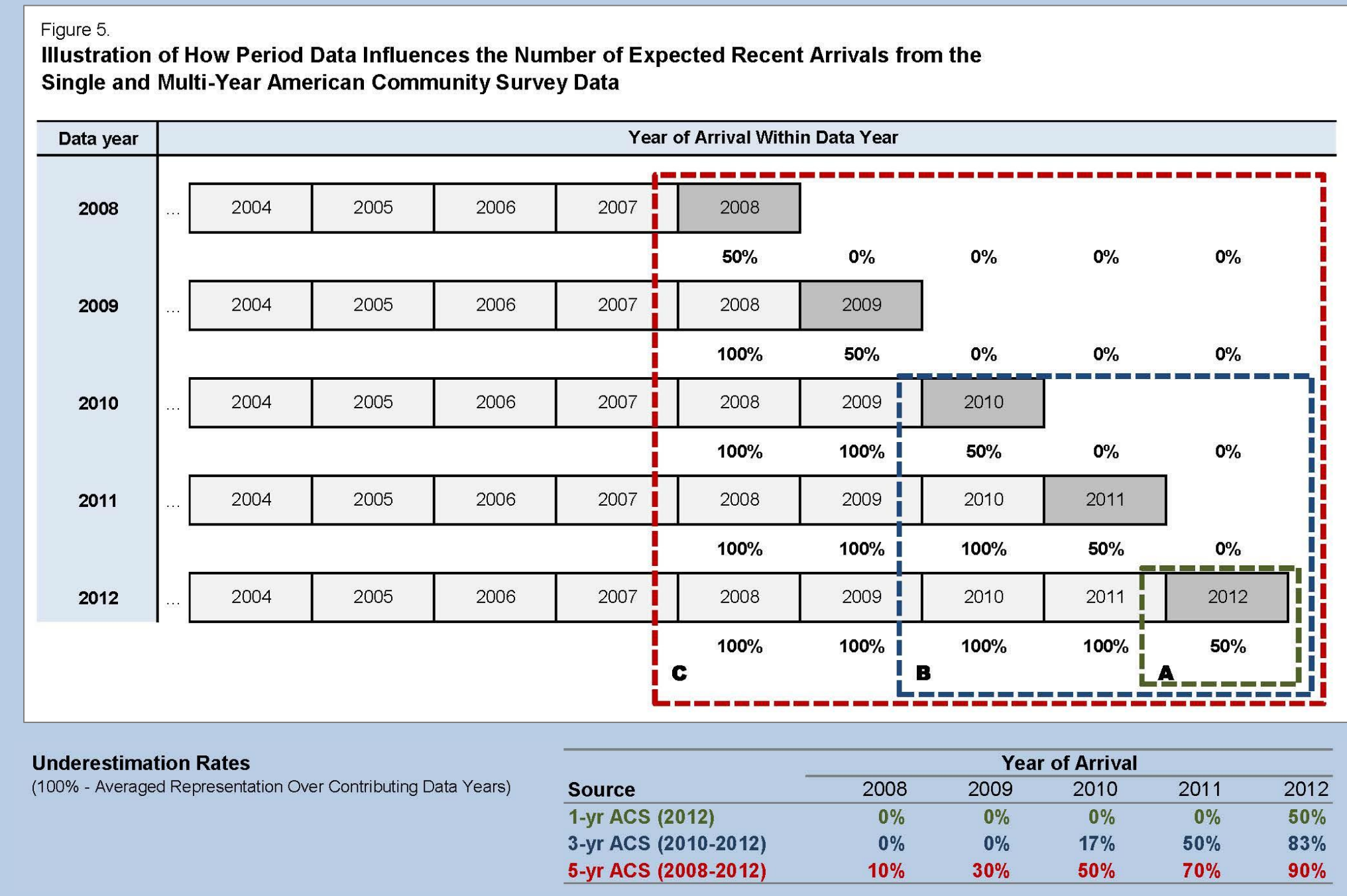Simply put: $p_{(Prior)} \approx 1$; $0 < p_{(Recent)} < 1$

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

$$\Pr(A) = \sum_{i=1}^{12} \Pr(A|m=i) * \Pr(m=i) \approx \frac{1}{12}\sum_{i=1}^{12} p_{(Recent)} = \frac{1}{12}\sum_{i=1}^{12}\frac{i}{12} = \frac{78}{144} \approx 0.5417$$
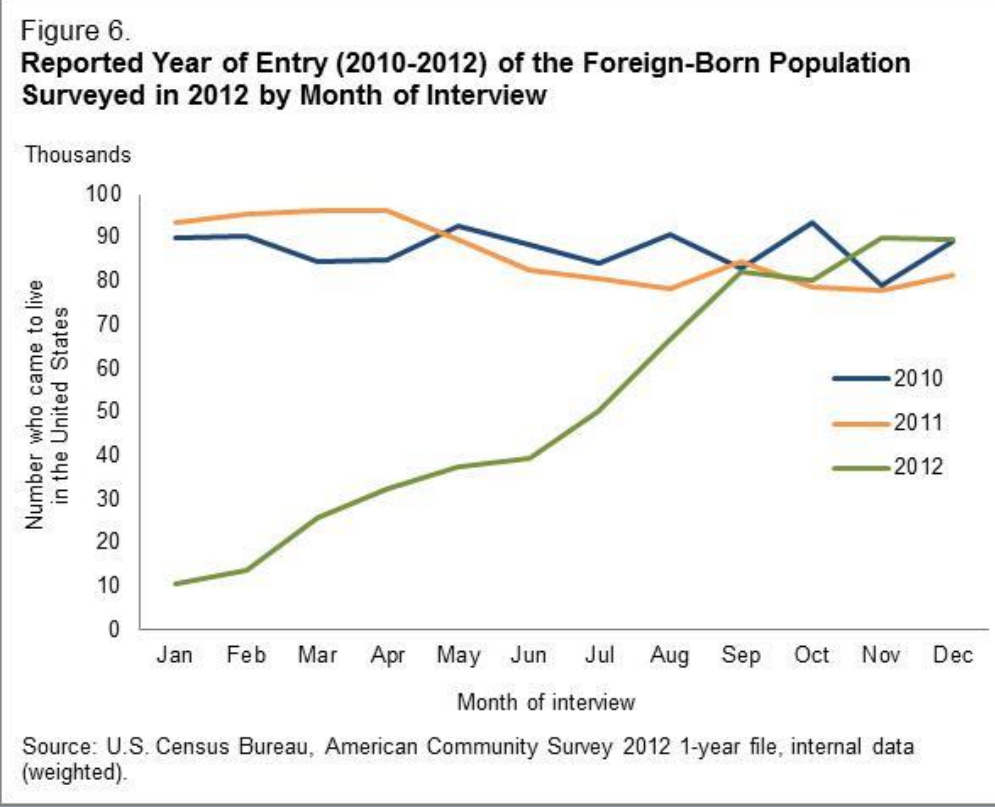
When adjusting for partial month values:

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

Recent immigrants whose housing units are interviewed in a given survey year have about a 50-50 chance to represented in the data. Based on the $\Pr(A) = 0.5$ finding, the estimated count of survey-year immigrant arrivals in the ACS has an expected value that is downward biased by about 50 percent.

Figure 5.
Illustration of How Period Data Influences the Number of Expected Recent Arrivals from the Single and Multi-Year American Community Survey Data

Data year	Year of Arrival Within Data Year								
2008	2004	2005	2006	2007	2008				
					50%	0%	0%	0%	0%
2009	2004	2005	2006	2007	2008	2009			
					100%	50%	0%	0%	0%
2010	2004	2005	2006	2007	2008	2009	2010		
					100%	100%	50%	0%	0%
2011	2004	2005	2006	2007	2008	2009	2010	2011	
					100%	100%	100%	50%	0%
2012	2004	2005	2006	2007	2008	2009	2010	2011	2012
					C	100%	B	100%	A
					100%	100%	100%	100%	50%

Underestimation Rates
(100% - Averaged Representation Over Contributing Data Years)

Source	Year of Arrival				
	2008	2009	2010	2011	2012
1-yr ACS (2012)	0%	0%	0%	0%	50%
3-yr ACS (2010-2012)	0%	0%	17%	50%	83%
5-yr ACS (2008-2012)	10%	30%	50%	70%	90%

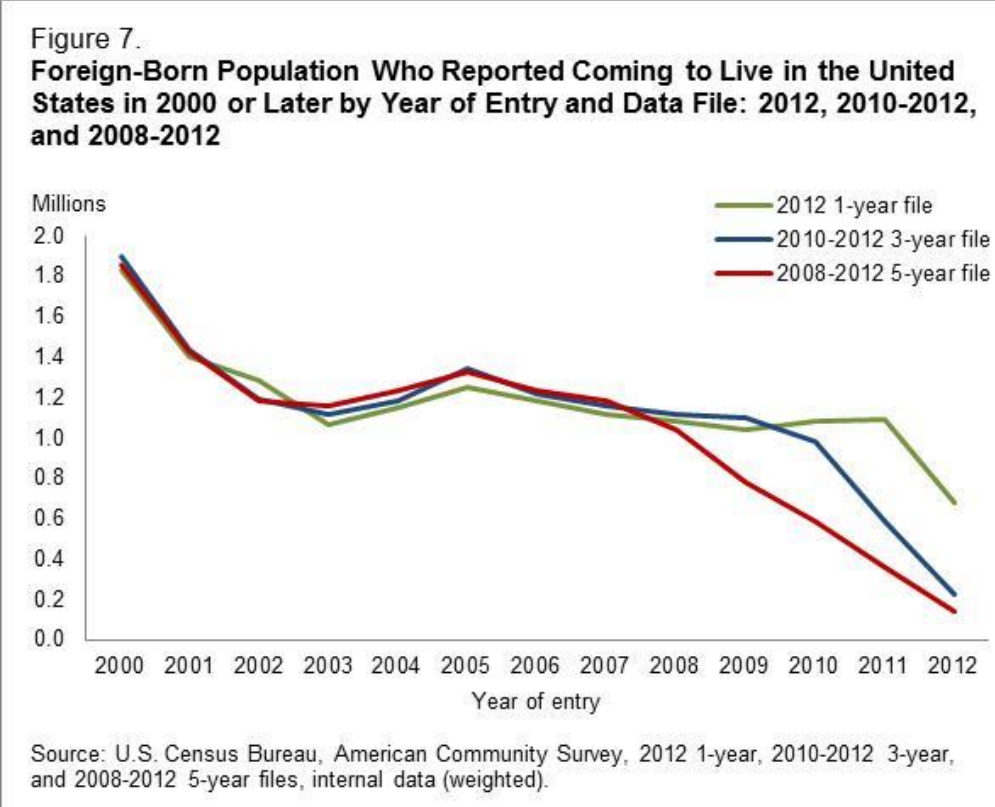## COMPOUNDED BIAS IN MULTI-YEAR FILES

The individual survey years that make up a multi-year file contribute data in roughly equal measure. For example, ACS 2010, 2011, and 2012 each contribute about the same to the 2010-2012 3-year file. The problem occurs when estimating counts for the most recent years of entry. For example, survey year 2010 does not have year of entry values of 2011, but it still contributed equally toward the year of entry estimate for 2011. Figure 5 illustrates this phenomena.

Figure 6.
Reported Year of Entry (2010-2012) of the Foreign-Born Population Surveyed in 2012 by Month of Interview

Source: U.S. Census Bureau, American Community Survey, 2012 1-year file, internal data (weighted)

RESULTS

The bias in estimates of survey-year arrivals can be illustrated using ACS single- and multi-year data. Using the 2012 1-year file, Figure 6 shows how immigrants who enter during the survey year have an initially low probability of inclusion that increases steadily with each passing month of data collection. This pattern is notably different for immigrants who reported arriving in earlier years.
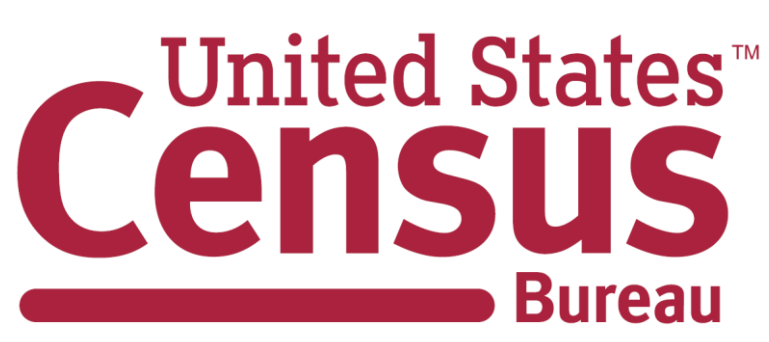
Using the 2012 1-year, 3-year, and 5-year files, Figure 7 shows how this bias is compounded in multi-year files, so that within each period, the underestimation is most severe in the most recent years.

Figure 7.
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and Data File: 2012, 2010-2012, and 2008-2012

Source: U.S. Census Bureau, American Community Survey, 2012 1-year, 2010-2012 3-year, and 2008-2012 5-year files, internal data (weighted).

CONCLUSION

The bias in the year-of-entry data, especially for the estimate of survey-year arrivals, is predominantly an artifact of the ACS data collection methodology rather than a coverage, response, or weighting issue, as has been suggested in the literature.

Given this bias in the estimates of survey-year arrivals, we urge researchers to use caution when interpreting the year-of-entry data, especially when drawing conclusions based on the number of "new" or "recent" immigrants, however that is defined.