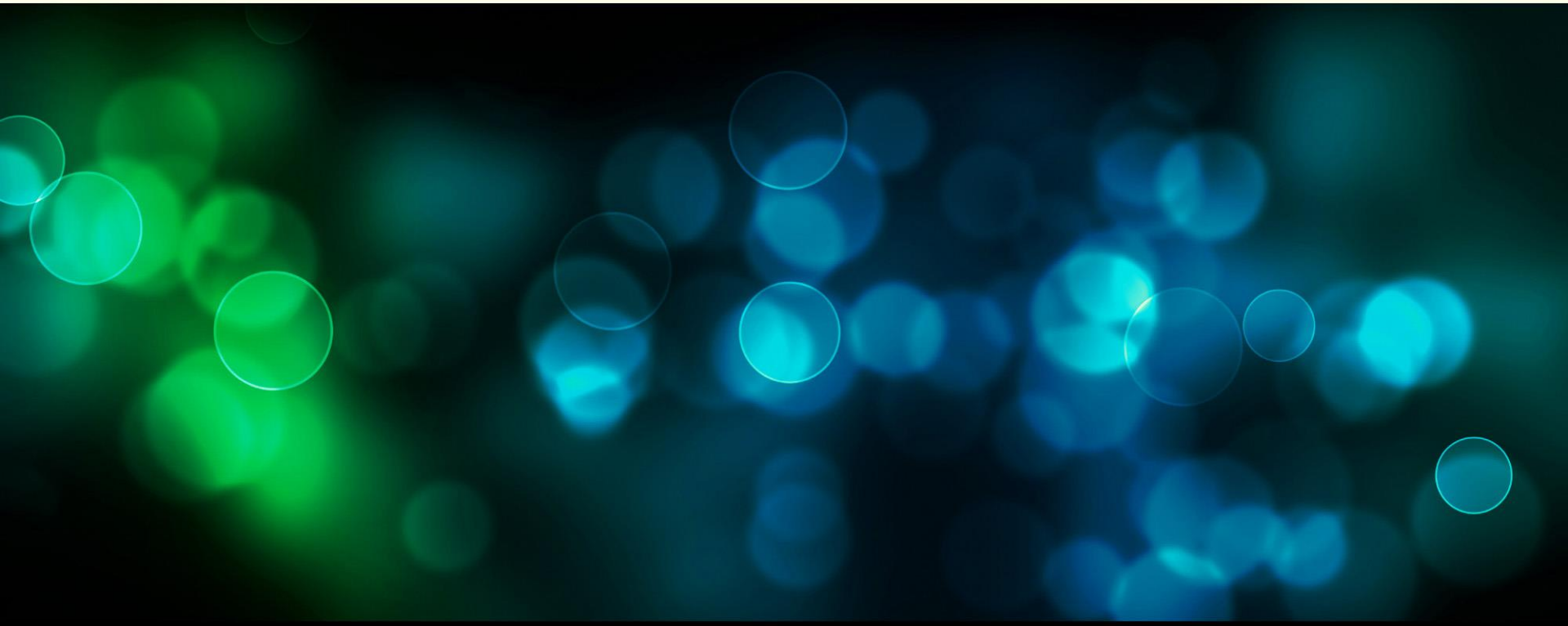
Aggregation of ACS 1-year Data into Multiyear Estimates



Presented to ACS Data Users Group

Dave Hubble, Westat

May 12, 2017

Overview

- Review of History & Current Situation
- Contributing Factors to Differences (i.e., limitations associated with aggregated 1-year estimates)
- New Supplemental Files
  - 2014 (and 2015) 1-year Estimates for 20,000+ Populations
  - 2010-2014 (and 2011-2015) Variance Replicate Estimates
- Recommendation for the development and release of a Tabular Aggregation System (TAS)

History

- In 2014, the production and release of official 3-year period estimates was discontinued.
- For those geographies that fell into the 20,000-65,000 population range, 5-year period estimates would now be the only data product available.
- A review of the loss of information lead data users to recommend the production of a supplemental set of 1-year period estimates that would include areas of 20,000 – 65,0000 population.
- The notion was that these 1-year estimates could be combined into pseudo 3-year estimates by using known aggregation techniques.

Current Situation

- The aggregation mechanics (documented in ACS Handbooks) have been demonstrated by users.
- Alicia VanOrman (PRB) gave an aggregation demonstration last fall using 1-year estimates from 2011 through 2013 from geographies of 65,000+.
- With noted differences, her aggregated pseudo 3-year estimates and MOEs did a fairly reasonable job of representing the corresponding official published 3-year (2011-2013) period estimates and MOEs.

Observation

- This result was generally as expected as the aggregation of 1-year estimates into multi-year estimates was the original plan for the ACS prior to development of the “period” estimate approach.
- Note that while the processes are fairly straightforward, they can be quite laborious.

But Why Are There Differences?

- Aren't the exact same set of interviews (and non-interviews) in both sets of estimates? Yes, but…
- For period estimates, all cases are "pooled" (i.e., processed\weighted together at the same time).
- For aggregate estimates, processed\weighted separately at different times and then joined.
- Of the various sources of differences this creates, can anything be done to mitigate them?
  - — Answers: No, Yes, and Maybe

What is it about Period Estimates that Contributes to Observed Differences?

- "Pooling" of data records allows for more refined non-response (NR) adjustment cells and corresponding NR factors.
- Current-Year "Vintage" applied along a few dimensions:
  - Population Controls - previous years adjusted to current year vintage (most sensitive when period crosses a census: Year ending in 0).
  - Geography - current definitions applied to all years in period (most sensitive when period crosses two years after a census: Year ending in 2).
  - Variable Definitions - standardized across years.
  - Inflation - previous years income and housing values adjusted to current year dollars (Compass Handbook - Appendix 5 provides guidance on use of "All Items CPI-U-RS Annual Averages" to adjust for this difference).

Other Factors beyond “Vintage”

- For multi-year period estimation only, a “model–assisted” weighting step to control to sub-county populations estimates is employed. 1-year estimates do not currently employ this step.
- For multi-year period variance estimation, a finite population correction factor (FPC) is applied to appropriately reduce estimates of variance to reflect the proportion of addresses in sample. Variance estimation for 1-year estimates do not currently employ this adjustment.
- With this expanded usage of 1-year estimates, possibly these estimation differences should be revisited.

Margin of Error (MOE)

- Recall that one of the limitations of these “custom” MOEs is that the covariance between the 1-year estimates is not accounted for in the approximation formulas.
- Interestingly, the observed differences in the MOEs were both positive and negative in VanOrman’s comparisons.
- One possible interpretation of this is that the potential impact of the FPC and covariance differences are small relative to the general variation in the 1-year MOE estimates being aggregated when compared to the 3-year period MOEs.
- Note, this also means the aggregated MOEs are not consistently conservative.

A Few Things to Note

- A more comprehensive assessment would come from aggregated 3-year estimates from geographic areas with 20,000-65,000 in population
  - Smaller sample sizes will contribute to greater deviations between 3-year period and aggregated estimates.
  - Differences from some of the contributing factors listed in the previous slides will be more sensitive within 20,000-65,000 population geographies relative to 65,000+ population geographies.
  - Especially true from the weighting and variance estimation differences related to sub-county estimates and high sampling rate geographies.

So, with that in mind…

- The first 1-year supplemental estimates were released for 2014 (7\16) and 2015 (10\16).
  - For geographies with 20,000+ population
  - Includes 23 geography levels: nation, state, county, place, metropolitan areas, congressional & school districts
  - Includes 58 high-level detailed tables
- A review of adequacy should be conducted.

Great Start, But…

- Beyond waiting for 2016 1-year, consider…
  - Generate & release corresponding 2012 and 2013 1-year supplemental estimates to avoid a two year gap in the 3-year series for 20,000-65,000 population geographies.
  - In fact, releasing 2011 1-year supplemental estimates would support repeating PRB's 3-year comparison efforts (period vs. aggregated), but now for the more directly relevant 20,000-65,000 population geographies.
  - Greater variation in estimates and MOEs likely to be observed in these smaller population geographies.

In Addition,…

- Spotted release of variance replicate estimates tables for 2010-14 (& 2011-15) 5-year estimates.
- Provided to allow users to calculate MOEs for their own aggregated estimates across geographies or categories within a table more accurately than from approximation methods in ACS documentation and employed in PBR work.
- So, knowing the intended use is to aggregate, I think there is merit in supplying this kind of table package for the supplemental 1-year estimates.

Looking Ahead…

- Possibly, all these supplemental estimates and ad hoc aggregation efforts could be made moot (or at least easier) by other approaches.
- In particular, the Micro-data Analysis System (MAS) comes to mind
  - Online remote access system to allow users to request custom tables from underlying micro-data
  - Could this be used? Possibly…

Micro-data Analysis System (MAS)

- However, MAS development has been very much challenged by desire to maximize utility while simultaneously maintaining confidentiality standards.
- The development of robust disclosure avoidance methods for the MAS advanced query system supporting customized geographies or variable categories without substantial table suppression has proven difficult.
- This outcome has been mostly driven by complexities associated with striking the right balance of the need to apply greater degrees of data perturbation to reduce suppression.

Tabular Aggregation System (TAS)?

- I wonder if refocusing the need of a tool (TAS) to just “aggregate published tables” would address a substantial part of “special tabulation” user needs?
- Use of published tables (with fixed geographies & categories) could overcome confidentiality concerns as each component in aggregation is already sufficient.
- Possibly, this approach would help in expediting development & release of a TAS to support the aggregation of the supplemental 1-year estimates, especially those in the 20,000-65,000 pop range.

Additional TAS Point on MOEs

- Option 1: Use the MOEs as published in each year or each geography that is to be aggregated. This has known limitations that are well documented.
- Option 2: Referring back to the release of supplemental variance replicate estimates files, their use in a TAS would result in more accurate MOEs associated with aggregate estimates.
- Hence the recommendation that they be produced for the supplemental 1-year estimates and used under Option 2 TAS to produce more robust aggregated MOEs.

Some Resources: Census Bureau

- ACS General Compass Handbook
  - https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf
- Design & Methodology Report - 2014
  - https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html
- Accuracy of the Data – PUMS 2011-2015
  - https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2011_2015AccuracyPUMS.pdf

Some Resources: Census Bureau (2)

- Table and Geography Changes
  - https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes.2015.html
- Supplemental Materials
  - https://www.census.gov/programs-surveys/acs/news/data-releases/2015/release.html
- The Schar, Freiman, and Lauger (2015), “Developing and Testing the Microdata Analysis System at the U.S. Census Bureau”
  - https://fcsm.sites.usa.gov/files/2016/03/J3_Schar_2015FCSM.pdf

Thanks for the opportunity.

DavidHubble@Westat.com