Incorporating Variance and Geographic Specificity into the Imputation Frame Used in Weighting the American Community Survey Group Quarters Sample

- Dirk Bullock, John M Jordan, Edward C Castro Jr.
- U.S. Census Bureau

- Disclaimer:
  - Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Outline:

- Motivation for research
- Group Quarters (GQ) Estimation in the American Community Survey (ACS)
- Collapsing algorithms and sampling variance for the frame
- GQ population CV results
- Summary/Conclusion

Motivation:
Frame Population Estimation and Variance Estimation

- We have a frame with all GQs eligible for the ACS.
- We update the populations on the frame with an adjustment factor.
- This helps determine areas where imputation is needed for coverage.
- Population updates to the frame are adjusted nationally.
  - A subnational adjustment will produce more accurate population updates and more accurate coverage.
- We currently do not account for the variance from the adjustment factor.

Motivation:
Research Questions

- What impact does a subnational frame adjustment have on the GQ population variances?
- What impact does accounting for the variance from the adjustment factor have on the GQ population variances?

Background:
Group Quarters

- A group quarters is a place where people live or stay, in a group living arrangement.
- GQs are classified into characteristic types defined as GQ Type Groups.
- Examples include: Prisons, Nursing Homes, College Dormitories, etc.
- GQs are sampled and weighted separately from housing units.

Background:
GQ Sampling

- GQ sampling is performed at the state level and each state has its own independent sampling rate.
- GQ persons are sampled based upon the size of the GQ where they live.
- When sampling, GQs are assigned size stratum based on population size.
- Small GQs are sampled as a whole.
- Large GQs are sampled in multiples of 10 people, proportional to the size of the GQ, at the state sampling rate.

Background:
Size Strata

- The estimation methodology is informed by the sampling methodology.
- Large GQs tend to be in sample more often than small GQs.
- The estimation size strata are the small GQs, the large GQs, and the GQs sampled with certainty.

Background:
Direct Methodology

- Prior to data year 2011, we used a direct estimation process.
- Direct Estimation
  - Inverse probability of selection
  - Non-interview adjustment
  - Coverage adjustment
- Substate estimates suffered from high variability.

Background:
Current Methodology

- Beginning in data year 2011, we impute whole GQ person records from GQ interviewed persons into GQs without interviews.
- We create a GQ imputation frame from the GQ sampling frame.
- GQs that are not sampled have their populations estimated by a Not-in-sample Adjustment Factor (NAF).
- We impute into all large GQs and we impute into small GQs to ensure substate coverage.
- We select whole person donors for imputation based on the geography, GQ Type, and Sex of the imputed GQ.

Background:
Current Methodology Continued

- Modified estimation procedure due to the inclusion of imputed GQ persons.
  - Modified inverse probability of selection
  - Use imputation frame to determine substate distribution
  - Coverage adjustment
- This improved the stability of our substate estimates.

Background:
Variance Estimation

- We use a replicate weighting methodology to estimate the variance through successive differences.
- The ACS uses 80 replicates.
- Weighting adjustments are made to both the production weight and the replicate weights.
- The variance accounts for imputations through design factors.

$$VAR(\hat{\theta}) = \frac{4}{80}\sum_{i=1}^{80}(\theta_i - \theta_0)^2$$

Research Methodology:
Collapsing Algorithms

- We have developed collapsing algorithms to estimate the NAF at localized levels of geography.
- Collapsing cells are defined as combinations of geography level, GQ Type Group, and GQ Size Stratum.
- No Collapsing means NAFs are calculated at the state geography.
- Complete Collapsing means NAFs are calculated at the national geography.

Research Methodology:
GQ Frame Variance

- Sampled GQs are assigned a baseweight equal to their inverse probability of selection at the state level.
- Replicate weighting methodology is applied to these sampled GQs to produce replicate weights.
- The NAF and its replicates are calculated from the sampled GQs and then applied to the GQs that are not in sample.
- Every GQ on the frame has a variance associated with it.
- There is variance associated with the frame adjustments during GQ estimation.

Results:
National CVs

Estimation Methodology	CVs
Direct	0.68%
Current	0.42%
No Collapsing	0.99%
Collapsing	0.97%
Complete Collapsing	0.97%

Data Source: 2016, 1-year ACS data.

CVs based on pre-controlled population totals.

DRB Approval Number: CBDRB-FY19-RAGLIN-B0014

Results:
State CVs

Estimation Methodology	Median CVs
Direct	4.64%
Current	2.98%
No Collapsing	4.67%
Collapsing	3.57%
Complete Collapsing	1.10%

Data Source: 2016, 1-year ACS data.

CVs based on pre-controlled population totals.

DRB Approval Number: CBDRB-FY19-RAGLIN-B0014

Results:
County CVs

Estimation Methodology	Median CVs
Direct	51.57%
Current	39.33%
No Collapsing	6.30%
Collapsing	5.25%
Complete Collapsing	1.81%

Data Source: 2016, 1-year ACS data.

CVs based on pre-controlled population totals.

DRB Approval Number: CBDRB-FY19-RAGLIN-B0014

Results:
Characteristic Variance

- Characteristics include: Sex, Age Group, Hispanic Origin, Race, Employment Status, Educational Attainment, and Poverty Status.
- The final population total CVs are consistent among the estimation methods.
- Often, the Direct method has the highest CVs.
- There is no discernable pattern between the collapsing parameters and the CVs.
- However, the No Collapsing Method usually has a larger CV compared to the Complete Collapsing Method.

Summary:

- We will begin to implement these methods for data year 2019.
- At the national level of geography, the Collapsing methods produce the largest CVs. However, the differences in CVs among all the methods are negligible.
- At the state level, the Collapsing methods are less than the Direct method, yet above the Current method in median CVs.
- At the county level, the Collapsing method is less than both the Direct method and the Current method in median CVs.

Further Research Performed:

- Collapsing patterns for different collapsing parameters.
  - Number of collapsed cells at each level of geography.
- Coverage rates for different collapsing parameters.
  - National, GQ Type Group, State, State by GQ Type Group.
- Characteristic estimates for different collapsing parameters.
  - Sex, Age Group, Hispanic Origin, Race, Employment Status, Educational Attainment, and Poverty Status.

Questions or Comments

- Dirk Bullock
- dirk.m.bullock@census.gov

References

- "Archived American Community Survey Design and Methodology Report: Chapter 11. Weighting and Estimation (Revised 12/2010)." U.S. Census Bureau, Dec. 2010. https://www.census.gov/content/dam/Census/library/publications/2010/acs/Chapter_11_RevisedDec2010.pdf. Accessed 18 Apr. 2019.
- "American Community Survey Design and Methodology Report: Chapter 11. Weighting and Estimation." U.S. Census Bureau, 30 Jan. 2014. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch11_2014.pdf? Accessed 04 Feb. 2019.
- "Archived American Community Survey Design and Methodology Report: Chapter 12. Variance Estimation (Revised 12/2010)." U.S. Census Bureau, Dec. 2010. https://www.census.gov/content/dam/Census/library/publications/2010/acs/Chapter_12_RevisedDec2010.pdf. Accessed 04 Feb. 2019.
- "American Community Survey Design and Methodology Report: Chapter 12. Variance Estimation." U.S. Census Bureau, 30 Jan. 2014. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch12_2014.pdf? Accessed 22 Apr. 2019.

References

- Asiala, M., Castro, E. (2013). Developing Replicate Weight-Based Methods to Account for Imputation Variance in Mass Imputation Application. https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Asiala_03.pdf. Accessed 04 Mar. 2019.
- Fay, R., & Train, G. (1995). Aspects of Survey and Model Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. Joint Statistical Meetings: Proceedings of the Section on Government Statistics (pp. 154-159). Alexandria, VA: American Statistical Association.
- "American Community Survey Design and Methodology Report: Chapter 4. Sample Design and Selection." U.S. Census Bureau, 30 Jan. 2014. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch04_2014.pdf?# Accessed 24 Apr. 2019.