Variance Reduction and Integration in the American Community Survey Three-Year Estimates

Michael Starsinic & Anthony Tersine
US Census Bureau

MYES Research Meeting
November 15, 2007

Outline

1. Brief Background
2. Two Primary Questions
   - Effects of “g-weighting”
   - “Typical” variance characteristics for small geographic levels
3. Integration
4. Conclusions

Background

- First multi-year ACS estimates produced for ACS 1999-2001/Census 2000 Long Form Comparison Study
- Tract-level variances higher, relative to LF, than expected (Van Auken et al 2005; Starsinic 2005)
- Primarily due to lack of subcounty controls for ACS estimates
  - LF used 100% Census counts

Background

- G-weight method devised to reduce variances for tracts (Fay 2005, 2006, 2007)

- G-weighting implemented for 3-year and 5-year estimates in ACS Multiyear Estimates Study (34 counties)
  - tracts targeted for 5-year
  - places/MCDs targeted for 3-year

- 5-year analysis (Starsinic & Tersine 2007) presented at JSM

Background

- 2003-2005 estimates were produced with and without g-weights (Internal)
- We can make direct comparisons between SEs for estimates with and without g-weights

Research Questions

- QUESTION 1: What effect does g-weighting have on variances of county, place, and MCD estimates?
- QUESTION 2: What are typical variance characteristics for three-year estimates of counties, places, and MCDs?

Methodology

- MYES data profiles – demographic, social, economic, housing
- 435 unique estimates grouped into 43 topics
  - e.g. age & sex, education, poverty rates, housing value, etc.

Methodology

- CV = $\frac{SE(Est)}{Est}$

- Analysis will focus on median CV ratios and medians CVs
  - CV(with g-weights) / CV(without g-weights) should be less than one if the g-weights decreased the variance

Number of Geographic Areas

- Areas that have a population of 20,000
- 30 counties
- 62 places
- 63 MCDs

Number of Geographic Areas

Pop Range	Counties	Places	MCDs
20,000-25,000	2	8	10
25,000-30,000	3	10	13
30,000-40,000	3	14	9
40,000-50,000	1	4	7
50,000-65,000	2	7	5
65,000-100,000	5	9	5
100,000-250,000	3	6	9
250,000+	11	4	5

Effects on the Estimates

- Difference in “with” vs. “without” g-weight estimates

	Number Estimates	| Z | > 1.645	% with | Z | > 1.645
County	13,050	251	1.9%
Place	26,970	654	2.4%
MCD	27,405	510	1.9%

Q1: G-Weight Results

Three-Year Median CV Ratios

Topic	County	Place	MCD
Total Housing Units	C	0.372	0.478
Total Population	C	0.634	0.695
Total Households	0.976	0.639	0.676
Age & Sex	0.997	0.823	0.878
Tenure	0.980	0.855	0.903
Occupied/Vacant	0.985	0.961	0.973
Race	0.999	0.976	0.984
Poverty	0.992	0.994	0.993
Ancestry	0.999	0.996	0.998

Q1: G-Weight Results

Three-Year Median CV Ratios – Places by Size of Area

Topic	20,000-25,000	25,000-30,000	50,000-65,000	65,000-100,000	100,000-250,000	250,000+
Total Housing Units	0.526	0.416	0.342	0.286	0.372	0.524
Total Population	0.628	0.609	0.593	0.651	0.666	0.709
Total Households	0.662	0.606	0.600	0.601	0.586	0.524
Age & Sex	0.795	0.811	0.816	0.831	0.851	0.837
Tenure	0.847	0.843	0.843	0.831	0.842	0.935
Occupied/Vacant	0.960	0.955	0.949	0.971	0.974	0.981
Race	0.966	0.957	0.991	0.982	0.990	0.985
Poverty	0.997	0.991	1.003	0.991	0.987	0.995
Ancestry	0.993	0.993	0.990	0.998	1.003	0.997

Q1: G-Weight Results

Three-Year Median CV Ratios – Places by Size of Estimate

Topic	< 100	100-250	250-500	500-1,000	1,000-2,000	2,000-5,000	5,000-10,000	10,000+
Total Housing Units							0.494	0.364
Total Population								0.634
Total Households							0.644	0.633
Age & Sex	0.999	0.987	0.959	0.915	0.913	0.885	0.849	0.731
Tenure				0.956	0.866	0.869	0.851	0.852
Occupied/Vacant	1.005	0.996	0.934	0.944	0.915	0.936	0.652	0.652
Race	1.003	1.000	0.985	0.982	0.955	0.927	0.859	0.732
Ancestry	1.000	1.002	0.998	0.993	1.003	0.985	0.987	0.986

Q1: G-Weight Results

Three-Year Median CV Ratios – MCDs by Size of Estimate

Topic	< 100	100-250	250-500	500-1,000	1,000-2,000	2,000-5,000	5,000-10,000	10,000+
Total Housing Units							0.830	0.660
Total Population								0.695
Total Households							0.830	0.660
Age & Sex	0.985	0.988	0.982	0.982	0.945	0.923	0.888	0.807
Tenure			0.926	0.966	0.939	0.904	0.903	0.886
Occupied/Vacant	0.993	0.944	0.965	0.960	0.976	0.956	0.857	0.683
Race	1.008	1.001	0.991	0.993	0.958	0.943	0.934	0.800
Ancestry	1.006	1.004	0.998	1.000	0.998	0.993	0.985	0.979

Q1: G-Weight Results

- Results consistent within counties
- Improvements were found in estimates of Total pop, HH, and HU values
- Larger estimates (in general) saw the most improvement

Transition

- Moving from g-weight analysis (Q1) to typical variance characteristics (Q2)
- Q1: Were looking at median ratios of CVs
- Q2: Now looking at median CVs
- CV > 61% means estimate not significantly different from zero at the 90% confidence level

Q2: Variance Characteristics

Three-Year Median CVs (Shown as percents)

Topic	County		Place		MCD	
	Without	With	Without	With	Without	With
Total Housing Units	C	C	2.7	0.9	2.5	1.1
Total Population	C	C	3.2	1.9	3.3	2.1
Total Households	0.8	0.8	2.6	1.5	2.6	1.5
Age & Sex	0.5	0.5	5.5	4.2	5.0	4.3
Tenure	2.4	2.3	4.8	3.9	4.4	3.7
Occupied/Vacant	8.0	8.0	22.9	22.1	24.7	23.7
Race	19.8	19.9	38.3	35.8	39.0	38.3
Poverty	11.0	11.0	23.6	23.7	21.8	21.6
Ancestry	17.8	17.8	30.6	30.4	28.1	28.1

Q2: Variance Characteristics

Three-Year Median CVs – Places by Size of Area with G-weights

Topic	20,000-25,000	25,000-30,000	50,000-65,000	65,000-100,000	100,000-250,000	250,000+
Total Housing Units	1.7	1.3	0.7	0.6	0.6	0.2
Total Population	2.8	2.7	1.9	1.6	1.5	0.3
Total Households	2.3	1.9	1.4	1.1	1.1	0.4
Age & Sex	7.5	6.3	4.5	3.6	3.4	0.8
Tenure	5.6	5.1	3.7	3.0	2.6	0.9
Occupied/Vacant	29.9	32.0	20.7	14.3	15.2	4.7
Race	43.4	44.3	40.1	32.7	27.8	13.4
Poverty	37.6	28.8	25.1	14.7	14.3	4.9
Ancestry	36.1	39.9	31.0	25.4	22.9	9.1

Q2: Variance Characteristics

Three-Year Median CVs – Places by Size of Estimate with G-Weights

Topic	< 100	100-250	250-500	500-1,000	1,000-2,000	2,000-5,000	5,000-10,000	10,000+
Total Housing Units							1.7	0.8
Total Population								1.9
Total Households							2.2	1.4
Age & Sex	50.9	34.2	21.6	15.2	12.4	8.8	5.7	2.3
Tenure				16.0	11.0	7.0	4.0	2.0
Occupied/Vacant	62.8	50.3	30.6	22.3	14.3	12.8	2.2	1.5
Race	94.3	54.9	39.7	29.6	20.9	15.4	9.7	2.6
Ancestry	73.7	45.6	34.8	25.1	17.9	12.6	8.2	6.0

Q2: Variance Characteristics

Three-Year Median CVs – MCDs by Size of Estimate with G-Weights

Topic	< 100	100-250	250-500	500-1,000	1,000-2,000	2,000-5,000	5,000-10,000	10,000+
Total Housing Units							2.8	1.0
Total Population								2.1
Total Households							2.8	1.4
Age & Sex	55.8	32.0	22.8	16.3	11.6	8.4	5.2	2.5
Tenure			20.9	16.0	12.0	7.3	3.5	2.2
Occupied/Vacant	69.9	51.8	33.8	23.5	15.1	9.8	3.1	1.4
Race	93.5	56.2	39.2	29.0	22.3	14.5	9.6	2.8
Ancestry	72.6	46.5	32.7	24.5	18.5	12.0	8.6	5.4

Integration – Decision Process

- Finalize variance improvements analysis
- Work with DID to have available the necessary administrative record data
- Work to fully specify the G-weighting module integrated with our production weighting system
- Have everything in place and tested before the 3-year weighting production in May 2008

Integration – Weighting

- Occurs in HU weighting prior to household person weighting
- Placed after the non-interview adjustments (MBF) but before the housing unit coverage adjustment (HPF).
- HU controls come after the G-weighting in the HU weighting.
- Person controls also come after in the household person weighting
- Group quarters population unaffected

Conclusions

- G-weights improve SEs for some key estimates at target level a lot
- Other estimates and other other levels show slight improvement or no change – nothing gets much worse (“do no harm”)
- Not much we can do about SEs for very small estimates

Contact Information

- If you have any questions or comments:

Michael.D.Starsinic@census.gov
Alfredo.Navarro@census.gov