Preliminary Investigation of Variance Issues Related to Generalized Regression Estimation used for American Community Survey Five Year Estimates

Richard A. Griffin and Keith A. Albright, U.S. Census Bureau

Key Words: Generalized Regression Model; Successive Difference Variance Estimator

The American Community Survey (ACS) multiyear housing unit (HU) weighting process includes a model-assisted step that uses a generalized regression model (GREG) to calibrate the HU weights using covariates derived from administrative records. The Census Bureau discovered that the variance estimates of some 5-year estimates at the national level may be larger than the corresponding 1-year estimates. The ACS standard error estimates are calculated using the successive difference variance estimator, which measures both sampling error and ratio estimation bias. One initial tool considered is calculating an alternative variance estimator targeting only sampling error. Some evidence exists that using this alternative variance estimator provides a possible explanation for this concern. This paper evaluates this concern using 2006–2010 ACS 5-year data.

This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

I. INTRODUCTION AND METHODOLOGICAL BACKGROUND

The ACS multiyear housing unit (HU) weighting process includes a model-assisted step that uses a generalized regression model (GREG or g-weighting) to calibrate the HU weights using covariates derived from administrative records. Use of this process began with the weighting of the first set of ACS multiyear data (2005–2007). Prior to implementation, the GREG process was tested in the Multiyear Estimates Study (MYES), which used 1999–2005 data from 34 of the 36 ACS test counties (U.S. Census Bureau, 2007). The MYES demonstrated the effect of GREG on reducing variances of estimates in sub county areas, particularly census tracts (Starsinic, 2007). However, preliminary Census Bureau work discovered that variances of some 5-year estimates have issues that we believe are related to the GREG procedure, issues that remain unresolved. For example, the standard error (SE) of the 5-year estimate for total occupied HUs in the United States is larger than the corresponding 1-year estimate. These variance estimates are calculated using the successive difference variance estimate calculated as follows:

$v(\hat{\theta}_0) = \frac{4}{80}\sum_{r=1}^{80}(\hat{\theta}_r - \hat{\theta}_0)^2\hat{\theta}_0)^2$, where $\hat{\theta}_0$ is the ACS production estimate and $\hat{\theta}_r$ is the estimate using the weights from replicate r.

This paper will attempt to determine if there are any underlying causes of these issues and if they can be eliminated or mitigated.

Some background useful to understanding the analysis follows.

Steps of the Implementation. The following five steps, outlined previously (Fay 2005a), form the basic elements of the GREG process for tract-level estimates:
1. Link administrative records to the ACS sampling frame (the MAF), dropping administrative records that cannot be linked.
2. Form unweighted tract-level totals of the linked administrative record characteristics.
3. Apply ACS sampling weights at the housing-unit level to the linked administrative record data that fall into the ACS sample. The weighted estimates at this step represent unbiased (or essentially unbiased) estimates of the unweighted totals in step 2.
4. Using GREG, calibrate the ACS sample weights so that the weighted administrative totals from the sample match the unweighted totals from step 2.
5. Use the new housing-unit weights in subsequent stages of the ACS weighting, which includes ratio and
raking/ratio estimation. Although the subsequent estimation steps adjust the new weights, the argument is that most of the variance reduction at the tract level will be retained in the final weights.

For GREG tract-level estimation seven age/sex categories and four race/ethnicity categories are used to define independent variables in the regression model.

The age/sex categories are as follows:

1. All persons age 0-17
2. All persons age 18-29
3. Males age 30-44
4. Females age 30-44
5. Males age 45-64
6. Females age 45-64
7. All persons age 65 and older

The race/ethnicity categories are as follows:

1. All Hispanics regardless of race
2. All Non-Hispanic Blacks
3. All Non-Hispanic Whites
4. All non-Hispanic Other races

The age/sex categories are potentially collapsed by checking two conditions without using race in the model: 1) is the regression equation solvable and 2) are all of the resulting weights greater than 0.5. If either condition fails then the age/sex categories are collapsed and the regression is attempted again. Two levels of collapsing are attempted:

1. Collapsing across age/sex into three categories: all persons age 0-17, all persons age 18-44, and all persons 45 and older.
2. Collapse all categories into a single cell of total administrative persons.

If the conditions still fail after the second level of collapsing, then administrative record data is not used. In this case, a one variable regression using only the frame count is attempted. If the conditions still fail then no GREG modeling is attempted.

If the regression passes using at least the single cell of total administrative persons, then an attempt is made to add race/ethnicity covariates as independent variables in the regression model. First, a collapsing procedure is run. The criteria for including a race/ethnicity category in the regression is that the administrative records universe count for the category being tested and the total for all other categories must be greater than 300 persons. This procedure is carried out first for the largest race/ethnicity category not including the non-Hispanic White category, then the next largest category and finally the smallest category. Three potential independent variables are added to the regression model based on the collapsing results:

1. Hispanics
2. Black non-Hispanics
3. Other Non-White Non-Hispanics

The following codes are used to indicate the model selection for each tract.

Codes for GREG Modelling

AGE/SEX Independent Variables		Ethnicity Independent Variables	
Choice Code	Model Choice	Rcell Code	Covariate in model
4	All seven categories	1	Hispanic
3	Three collapsed categories	2	Black Non-Hispanics
2	One total population category	3	Other Non-White Non-Hispanics
1	Only a frame count	0	No Ethnicity variable in model
0	No GREG modelling		

For ethnicity, the code for each variable included in the model selected for the tract is used to create a one, two, or three digit code indicating the ethnicity variables in the model. For example, 23 indicates that only the Black Non-Hispanics and Other Non-White Non-Hispanic independent variables were used (Hispanic variable omitted). 123, indicates all three potential ethnicity independent variables were used in the model.

The calibrated GREG weights are then further adjusted with HU and person post-stratification adjustments to obtain final weights. Section II describes the research questions documented in Albright (2013) and presents results and analysis for each question. Section III summarizes the analysis of the major issues and presents some potential future research.

II. RESEARCH QUESTIONS, METHODOLOGY, RESULTS AND ANALYSIS

The specific research questions presented in this section were motivated by things Census Bureau staff observed regarding the variances of 5-year estimates, as well as the GREG adjustments that were applied to individual sample records. For example, the 5-year estimate of occupied HUs in the United States has a variance estimate that is larger than the corresponding 1-year estimate, and its replicate estimates are all larger than the point estimate. These two results have also been observed in other estimates. This report will attempt to determine what causes this issue. Answers to these questions may give insight into changes that could be made to the GREG process and variance computation to mitigate or eliminate these issues.

Our analysis uses the 2006–2010 ACS data. All the data in all the tables and text of this report uses the production 2006-2010 ACS data. Much of the methodology involves comparing estimates produced using GREG (the original estimates) with estimates produced without GREG. Although a significant portion of the analysis focuses on census tracts, we also include other summary levels as well. Throughout the paper the term "summary levels" refers to national, state, county, Minor Civil Division (MCD), place, and tract, unless otherwise specified.

A. For what estimates are variances of 5-year national level estimates larger than variances of the corresponding 1-year estimates?

We noted earlier that this was observed for the estimate of occupied housing units in the United States. This phenomenon was also observed, to a lesser extent, at the state level. We looked at this estimate using the 2006–20010 data that was weighted without GREG and using the original GREG weights with a modified variance formula. In both cases, the SE for this estimate was much lower than the published SE. This portion of the research is intended to find out the extent that this issue exists for other estimates and summary levels and whether it is more likely to occur with the use of GREG.

A profile line refers to a specific ACS estimate at a summary level. For example at the state level one profile line is "Male householder, no wife present, family". This is the state level estimate of the number of families with a male householder with no wife living with the family. Profile lines are grouped into topic categories. For example, topic Educational Attainment at the national level consists of 7 estimates related to the educational attainment of persons. At the state level there are 51x7 = 351 of these estimates. We have identified 5-year profile lines where the SE is at least 90% of the corresponding 1-year SE. Frequency distributions have been be run to determine if there are specific types of estimates where this tends to occur and at what summary levels.

We have also recalculated standard errors using a slightly different alternative variance formula.

$$Var(Y) = \frac{4}{80}\sum_{i=1}^{80}\left(Y_i - Y_{avg}\right)^2 \quad (1a)$$

where Y_i is the i^{th} replicate estimate and Y_{avg} is the average of the 80 replicate estimates. This variance estimator differs slightly from the variance formula currently used by the ACS, in which the squared term uses the point estimate Y_0 (also called the full sample estimate) instead of Y_{avg} :

$$Var(Y) = \frac{4}{80}\sum_{i=1}^{80}(Y_i - Y_0)^2 \quad (1b)$$

We then compared the recalculated standard errors to the original values. Typically, Y_0 and Y_{avg} are close and there is little difference between (1a) and (1b). However, for the situation addressed in this research question, it could be that $(Y_i - Y_{avg})$ is smaller than $(Y_i - Y_0)$ for most, if not all, replicate estimates. This would cause a large difference between (1a) and (1b).

1. **Results**

For both 5-year estimates using GREG and 5-year estimates computed without using GREG (NOGREG) , the number of estimates for each topic and the number of those estimates with a SE ratio of 5-year to 1-year > 0.9 was calculated.

Tables A1 (Production variance estimate) and A2 (Alternative variance estimate) provide the proportion of estimates with a SE ratio greater than 0.9 and the unweighted average CV ratio over topics for both GREG and NOGREG. Note that based only on sample size considerations the sampling error only CV ratio would be expected to be close to $\sqrt{0.2} = 0.447$ so that those ratios larger than 0.9 are clearly a concern.

These tables make it very clear that the concern about a high CV ratio for GREG estimates is limited to the national geographic level using the production variance estimate.

From Table A1 (Production variance estimate), using GREG, 23.28% (88/378) of national estimates have a SE ratio greater than 0.9. The unweighted average CV ratio (5yr/1yr) over topics is 0.773. For NOGREG estimation, 0.26% (1/378) national estimates have a SE ratio greater than 0.9 and the unweighted average CV ratio (5yr/1yr) over topics is 0.465. The other geographic areas show small differences between GREG and NOGREG with less than 1 percent of the SE ratios greater than 0.9 and an unweighted average CV ratio between 0.460 and 0.491.

From Table A2 (Alternative variance estimate), there is little difference between GREG and NOGREG for all geographic areas including the nation. At the national level there were zero estimates out of 378 with a SE ratio greater than 0.9 for both GREG (average CV ratio = 0.43)and NOGREG (average CV ratio = 0.44). The other geographic areas show small differences between GREG and NOGREG with less than 1 percent of the SE ratios greater than 0.9 and an unweighted average CV ratio between 0.438 and 0.493.

Table A1		Production Variance Estimation		
Summary	% SE Ratio	Mean CV	% SE Ratio	Mean CV
Level	>.9 GREG	Ratio GREG	>.9 NOGREG	Ratio NOGREG
Nation	23.28%	0.773	0.26%	0.465
State	0.44%	0.473	0.27%	0.460
County	0.81%	0.475	0.84%	0.483
MCDMCD	0.84%	0.469	0.83%	0.489
Place	0.81%	0.473	0.82%	0.491

Table A2	Alternative Variance Estimation			
Summary	% CV Ratio	Mean CV	% CV Ratio	Mean CV
Level	>.9 GREG	Ratio GREG	>.9 NOGREG	Ratio NOGREG
Nation	0.00%	0.43	0.00%	0.440
State	0.24%	0.438	0.26%	0.450
County	0.80%	0.477	0.83%	0.487
MCD	0.87%	0.467	0.86%	0.487
Place	0.83%	0.473	0.83%	0.493

2. Analysis

According to Fay (2007), the ACS use of GREG at the tract-level results in "indirectly estimating characteristics at higher levels through summation. Thus, small biases at the weighting area level potentially can aggregate to high levels where, although relatively small, they grow in relation to the sampling variance." This could well be a major contributor to the concern for observed results at the national level when comparing ACS 5-year estimates with 1-year estimates. The successive difference variance estimator used for ACS (see Fay and Train 1995) is a replication variance estimator. Replication variance estimation is further discussed in Fay (1984). An instructive paper providing simulations of replication variance estimation is given by Judkins (1990).

None of these references suggests a variance estimator limited to the sum of squared differences about the replicate mean. They are based on "taking the mean square difference among the replicate estimates as the variance estimate" (Judkins 1990). The mean square difference is estimated by summing the squared differences between the replicate estimates and the full sample estimate. Thus the full sample estimate is treated as the "true value target" for obtaining the mean square error. The successive difference variance estimate can be written as follows:

$v(\hat{\theta}_0) = \frac{4}{80}\sum_{r=1}^{80}(\hat{\theta}_r - \hat{\theta}_0)^2 = \frac{4}{80}[\sum_{r=1}^{80}(\hat{\theta}_r - \overline{\theta}_r)^2 + 80(\overline{\theta}_r - \hat{\theta}_0)^2]$, where $\hat{\theta}_0$ is the ACS production estimate and $\hat{\theta}_r$ is the estimate using the weights from replicate r.

This variance estimate is thus constructed as the sum of a term for the variance among the replicates and a term for the squared difference between the mean of the replicate estimates and the full sample estimate. This second term is a first order approximation of the bias due to ratio estimation (verified via personal correspondence with Robert Fay who developed the successive difference variance estimator). Variance estimation using replication and in particular the ACS successive difference variance estimator is better described as a mean squared error estimator. However only the ratio estimation bias is measured, other forms of bias such as response bias and missing data are not measured by the successive difference estimator. This makes it seem that forming the variance estimator as only the variance of replicate estimates about their mean provides estimates of sampling variance but not mean squared error. If the squared bias second term is meaningful then it may be best to include it in an estimate of mean squared error. This is done by using the production variance estimate.

The results described above indicate that the concern for variances of 5-year national level estimates larger than variances of the corresponding 1-year estimates is more properly expressed as a concern about **mean squared error (sampling error and ratio estimation bias only)** of 5-year national level estimates that are larger than the **mean squared error** of the corresponding 1-year estimates. This concern only applies at the national level and appears to be caused by small biases due to GREG at the weighting area (tract) level aggregating to high levels where, although relatively small, they grow in relation to the sampling variance. Estimates formed without using GREG do not share this national level concern. Thus, this phenomenon at the national level is the result of ratio estimation bias and not sampling variance.

For both GREG and NOGREG estimates the production variance estimator, which actually is an estimate of mean squared error (ratio estimation bias only), is appropriate. The NOGREG estimate likely has less ratio estimation bias but is not unbiased.

This analysis is discussed in more detail in Griffin (2014) which provides an example using a simplified GREG procedure and assuming simple random sampling.

B. Do some geographies have a disproportionate effect on national level variances?

This question ties into question A, regarding the increased 5-year variances, and may provide an explanation for that issue. We identify profile lines where the SE of the national level estimate is higher than the corresponding standard error that results from not using GREG. Then we try to determine if these higher variances are being driven by higher variances at lower levels of geography or if the problem is systematic. For instance, for a particular profile line, estimates in some tracts may have much higher variance resulting from the use of GREG than they would otherwise (we noted earlier that the GREG process produced lower variances for tract-level estimates, but this is only in a general sense, not universally true for all estimates in all tracts).

1. Results

Table B1 provides the overall mean standard error ratios as well as the number of profile lines (N), standard deviation (SD), minimum, 99th percentile and maximum for each geographic area summary level.

Table B1 Summary Statistics for Distribution of Standard Error Ratios (GREG/NOGREG)

Summary Level	N	Mean	SD	Minimum	99th Percentile	Maximum
Nation	380	1.681	0.983	0.776	5.820	7.404
State	19,280	1.022	0.193	0.319	1.627	14.589
County	1,098,260	0.994	0.540	0	1.368	152.542
Minor civil division	5,401,908	0.991	0.153	0.092	1.454	10.366
Place	5,447,605	1.011	0.170	0	1.586	17.217
Tract	21,710,011	0.944	0.213	0	1.544	17.217

Table B2 provides the topic for the profile line with the maximum standard error ratio shown in Table B1 for each geographic summary level. The mean of the standard error ratio over the profile lines for that topic and the national summary level mean for that topic are also provided. This information is used in section B.2 below to analyze the effect of outliers on aggregated national-level estimates.

Table B2 Topics with highest Standard Error Ratios (GREG/NOGREG)

Summary Level	Topic with Maximum Ratio	Mean Ratio for this Topic	National-Level Mean Ratio for Topic
Nation	Household Type (HHTYPE)	3.641	3.641
State	Hispanic Origin Status (HISP)	1.104	1.096
County	HISP	1.108	1.096
Minor civil division	Year of Entry (YOE)	1.005	1.446
Place	Class of worker (COW)	0.943	1.425
Tract	COW	1.019	1.425

2. Analysis

We want to determine if some geographic summary levels have a disproportionate effect on national estimates. Do outliers at some summary level aggregate to the national-level producing the large variance estimates for some 5-year estimates? How do the extreme outliers compare to the 99th percentile outliers? Table B1 shows that the maximum standard error ratios are extreme. For states and minor civil divisions the maximum is about 9 times larger than the 99th percentile ratio. For places and tracts the maximum is about 11 times larger. For counties, the maximum is exceptionally extreme at over 100 times larger than the 99th percentile ratio. As shown in Table B2 there are some differences across summary levels in the topic producing the maximum ratio. HHTYPE (household type) is the topic with the highest ratio at the national-level. However, for the State and County levels, the topic HISP (Hispanic) had the highest ratio, the highest ratio for minor civil divisions is the topic YOE (year of entry), and the highest ratio for places and tracts is for the topic COW (class of worker). For each of these summary levels, the mean ratio over all the profile lines for that topic was not far from one indicating that the topic in general does not produce extreme standard errors using GREG. In addition at the national level, these same topics did not have particularly large ratios. For example, the topic COW for tracts has a maximum ratio of 17.217. However, the mean ratio for this topic at the tract level is 1.019 and the mean ratio at the national level for topic COW is 1.425 compared with the average mean ratio of 1.786 over topics at the national-level. If topic COW had a disproportionate effect on national level variances, the mean ratio at the tract level would be large and national level mean ratio would be significantly larger than the average mean ratio over topics. Thus, there is no indication from this data that any particular profile lines, topics, or geographies are having a disproportionate effect on national-level variances.

These standard error ratios use the production variance estimator and are more properly called **root mean squared error ratios (ratio estimation bias only)**. The number of profile lines for counties, minor civil divisions, places and tracts is in the millions. The mean reduction in root mean squared error ratio using GREG is 0.6% for counties, 0.9% for minor civil divisions, -1.1% (an average increase) for places, and 5.6% for tracts. For states, there is a mean increase of 2.2% and the mean increase for the nation is 68.1%. As discussed in section A.2, small biases due to GREG at the tract-level can aggregate to high levels where, although relatively small, they grow in relation to the sampling variance. This is the likely reason for these ratios averaging an increase in root mean squared error for larger geographic summary levels. The sampling variance portion of the mean squared error is likely similar for GREG and NOGREG for larger summary levels. Since GREG improves estimation at the tract level and it is necessary for production purposes to only have one set of weights, it is not feasible to use GREG based weights for smaller levels of geography and NOGREG weights for larger geographic areas.

C. What characteristics have a point estimate that is higher (lower) than all its replicate estimates?

In any estimate, the average of the 80 replicate estimates should be close to the full sample estimate (the same is true for replicate weights of sample cases). However, we have found instances where all replicate estimates are either higher or lower than the full sample estimate. This situation may have a detrimental effect on estimated variances and be related to the issue that is the subject of question A.. Based on the decomposition of the production variance estimate discussed in III.A.2, this is an indication that the bias is more of a concern than the variance for an estimate. All replicates either higher or lower than the full sample estimate will make $\overline{\theta}_r$ more different than $\hat{\theta}_0$ increasing the bias portion of the decomposition. ; higher replicate differences are not cancelled by lower replicate differences. We wish to determine which characteristics this happens with and at what summary levels and then determine if there are any patterns or trends that may indicate why it happens. We also will determine if it happens more often when using GREG than without GREG.

First we identify 5-year profile lines where all the replicate estimates are higher (lower) than the full sample estimate. Then run frequency distributions to find what types of estimates this tends to occur and at what summary levels. The same frequency distributions are then run using the non-GREG versions of these estimates and the results are compared.

1. Results

For each topic for each of the geographic areas (summary level): nation, state, county, minor civil division, place, and tract, the number of estimates and the number of these estimates for which the point estimate is higher or lower than all its replicate estimates were calculated. Table C1 provides summary results across topics at the summary area level.

Table C1 Point Estimates Higher or Lower than all Replicate Estimates

	Number		All Reps	Percent all reps	
.	Of		Hi or Low	Hi or Low	
Summary Area	Estimates	GREG	NOGREG	GREG	NOGREG
Nation	378	75	1	19.841%	0.265%
State	19245	25	3	0.130%	0.016%
County	1092179	189	161	0.017%	0.015%
Minor civil division	5360077	711	645	0.013%	0.012%
Place	5408782	259	213	0.005%	0.004%
Tract	21566182	269	227	0.001%	0.001%

2. Analysis

From Table C1, for the national level estimates using GREG, about 20 the percent of estimates have all replicate estimates higher or lower than the point estimate . All the NOGREG estimates as well as the GREG estimates for sub-national levels have well less than 1% of estimates with all replicate estimates higher or lower than the point estimate. Thus all replicate estimate higher or lower than the point estimate is only notable for GREG estimates at the nation-level.

Table C2 shows Topics that are notable in terms of the number of profile lines with all the replicate estimates higher or lower than the point estimate.

Table C2 Some Topics with Large Proportion of All high or low Replicate Estimates

Topic Description	Number of Profile Lines	Number of Profile Lines with all Replicates Hi or Low
Mortgage	2	2
Occupied housing Units	1	1
Only one race	1	1
Tenure	2	2
Total Households	1	1
Total Housing Units	1	1
Educational Attainment	7	7
Household Type	13	8
Migration	7	4
Relationship	7	3
Housing value	8	5

The correlation between the topic proportion of all profile lines with replicate estimates higher or lower than the point estimate and the topic mean standard error ratio GREG/NOGREG is 0.482 so that a simple regression with the SE ratio as the dependent variable (y) and the proportion of all replicate estimates higher or lower than the point estimate as the independent variable (x) has a R^2 of 0.232 (estimated regression equation $y = 1.466 + 1.145x$). A correlation of 0.482 indicates a modest or moderate positive relationship (Taylor 1990).

All replicates estimates higher or lower than the point estimate is an indication of ratio estimation bias. These results are consistent with the previous results and indicate that it is only at the national level using GREG that there is a substantial bias component likely caused by aggregation.

The alternative variance estimator is a measure of sampling error and does not include ratio estimation bias. Ratio estimation bias exists for both the GREG and NOGREG estimates and is likely larger for GREG estimates. The production variance estimator measures the sum of the sampling variance and the square of the ratio estimation bias. Denote production variance estimate as $\hat{V}_P$ and the alternative variance estimate as $\hat{V}_A$. Also denote the absolute value of the ratio estimation bias of an estimate as B. Then

$\hat{V}_P = \hat{V}_A + B^2$ and $B = \sqrt{\hat{V}_p - \hat{V}_A}$. We use this equation for each profile line for each topic at the national level (378 profile lines). There are a total of 50 profiles lines for the topics in Table C2 with a high number of profile lines with all replicates higher or lower than the point estimate. Table C3 shows average coefficients of variation for GREG and NOGREG estimates using the alternative variance estimator and the production variance estimator for all 378 profile lines and for the 50 profile lines for the topics in Table C2. The relative absolute bias using $B = \sqrt{\hat{V}_p - \hat{V}_A}$ is also shown.

Table C3 Average Statistics Over 378 Profile Lines and Over 50 Profile Lines

Statistic	Averages Over all 378 Profile Lines	Averages Over 50 Profile Lines from Table C2
Alternative Variance Estimator GREG CV	0.226	0.0958
Alternative Variance Estimator NOGREG CV	0.235	0.0996
Production Variance Estimator GREG CV	0.317	0.293
Production Variance Estimator NOGREG CV	0.243	0.117
Relative Absolute Ratio Estimation Bias GREG	0.00178	0.00265
Relative Absolute Ratio Estimation Bias NOGREG	0.000436	0.000528

Looking at Table C3, the average CV based on the alternative variance estimator, which measures sampling error, is about the same for GREG and NOGREG over all 378 profile lines and over the 50 profile lines representing the topics from Table C2 with a high proportion of all high or low profile lines. The average CV based on the production variance estimator, which measures a mean squared error (ratio estimation bias only), is greater for GREG than NOGREG over all 378 profile lines as well as the 50 profile lines. GREG estimates have more ratio estimation bias, **as measured by** $B = \sqrt{\hat{V}_p - \hat{V}_A}$, than NOGREG estimates.

Comparing the 50 profile lines column with the 378 profile line column, the topics with a high proportion of all high or low profile lines have lower sampling error as measured by the alternative variance estimator. They also have slightly larger average relative absolute ratio estimation bias. The level of sampling error is not related to the bias of estimates. Using $B = \sqrt{\hat{V}_p - \hat{V}_A}$ to estimate ratio bias may have limitations. The second term of the production variance estimator is a first order approximation of the bias due to ratio estimation. This does not mean that it is an unbiased or stable estimator of ratio estimations bias. This second term is $4(\overline{\theta}_r - \hat{\theta}_0)^2$. When an estimate for a profile estimate line has all replicate estimates higher or lower than the point estimate this second term is greater since all θ_r in $\overline{\theta}_r$ are higher (or lower) than the point estimate. There is no positive difference compensated by a negative difference in calculating $\overline{\theta}_r$. This fact combined with limitations in the first order approximation estimator could well be a contributor to the increased **estimated** average absolute relative ratio bias for the topics with a high proportion of profile lines with all replicate estimates higher or lower than the point estimate.

The variance at the national level is very small for both GREGH and NOGREG but the bias, although small in a relative sense, is larger for GREG than NOGREG.

III. SUMMARY AND POTENTIAL FUTURE RESEARCH

The concern for variances of 5-year national level estimates larger than variances of the corresponding 1-year estimates is more properly expressed as a concern about **mean squared error (ratio estimation bias only)** of 5-year national level estimates that are larger than the **mean squared error** of the corresponding 1-year estimates. This concern only applies at the national level and appears to be caused by small ratio estimation biases due to GREG at the weighting area (tract) level aggregating to high levels where, although relatively small, they grow in relation to the sampling variance. Estimates formed without using GREG do not share this national level concern. Thus, this phenomenon at the national level is the result of ratio estimation bias and not sampling variance. For both GREG and NOGREG estimates the production variance estimator, which actually is an estimate of mean squared error (ratio estimation bias only), is appropriate. The NOGREG estimate likely has less bias but is not unbiased.

The sampling variance portion of the mean squared error is likely similar for GREG and NOGREG.

All replicates estimates higher or lower than the point estimate is an indication of ratio estimation bias, as measured by the production variance estimator, in the point estimates themselves. The second term of the decomposition of the production variance estimate is $4(\overline{\theta}_r - \hat{\theta}_0)^2$**.** When all replicates are larger or smaller than the production estimate the absolute value of the difference between the replicate average and the production estimate is greater than if some replicate estimates are higher and some are lower than the production estimate. Limitations in the production variance estimator's accuracy in estimation of the ratio estimation bias of mean squared error could well be a contributor to the increased ratio bias measured.

These results indicate that it is only at the national level using GREG that there is a substantial ratio estimation bias component, likely caused by aggregation. The variance at the national level is very small for both GREGH and NOGREG but the bias, although small in a relative sense, is larger for GREG than NOGREG. The GREG weighting does not result in large changes to the point estimates. The advantage in using GREG is reduction in variances of the point estimates at the tract level.

For all results presented here the form of the GREG model is determined using the full sample weight and that same model is used for each replicate for variance estimation. We had planned to reweight allowing the GREG model to be selected independently for each replicate. Due to resource limitations, we have not done this for this report. This may be done in the future, if after reviewing all the results presented, we think that doing so would provide useful information and resources are available.

Finally, none of the analyses in this report suggest a need for changes in the production implementation of GREG or to the production variance estimation.

IV. REFERENCES

Fay, R. (1984), "Theory and Application of Replicate Weighting for Variance Calculations," Proceedings of the Survey Research Section the American Statistical Association .

Fay, R. (1989), "Theory and Application of Replicate Weighting for Variance Calculation," Proceedings of the Survey Research Section of the American Statistical Association.

Fay, R. and Train, G. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," Proceedings of the Survey Research Section of the American Statistical Association.

Fay, R.E. (2007) "Imbedding Model-Assisted Estimation into ACS: The Impact on Users," paper presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA, 5 Nov.-7 Nov. 2007, available from http://www.fcsm.gov/07papers/Fay.VII-B.pdf

Griffin, R. (2014), "Preliminary Investigation of Variance Issues Related to Generalized Regression Estimation," 2014 American Community Survey Research Memorandum Series, ACS-R-draft

Judkins, D. (1990), "Fay's Method of Variance Estimation," Journal of Official Statistics, Vol. 6 No. 3, pp223-239.

Starsinic, M. and Tersine, A .(2007), " Analysis of Variance Estimates from American Community Survey Multiyear Estimates," *Proceedings of the Joint Statistical Meetings on CD-ROM,* American Statistical Association, pp. 3011-3017

Taylor, R. (1990), "Interpretation of the Correlation Coefficient: A Basic Review," JDMS 1:35-39 Jan/Feb.

U.S. Census Bureau (2007), "Accuracy of the Data (Multiyear Estimates Study)", http://www.census.gov/acs/www/Downloads/methodology/special_data_studies/multiyear_estimates/Accuracy_of_the_Data_MYE.pdf