A Simple Solution For Adjusting Weights in a Complex Sample

Quatracia Lucky

Yeng Xiong, Ph.D.

Acknowledgements

Jacklyn Jonas

Jacob Sandruck

Daniel Whitehead

Outline

- Survey background

- Research objective/requirements

- Considerations

- Simulation Study and recommendations

- Conclusion

Annual Integrated Economic Survey (AIES)

- Newly developed survey integrating seven industry-specific surveys to measure the economic activity in the US at the national and varying subnational levels
- Covers a wide range of economic industries, which are determined by the North American Industry Classification System (NAICS)
  - National estimates at NAICS6 level
    - 445210 – Meat Markets
    - 445220 – Fish and Seafood Markets
    - 445291 – Baked Good Stores
  - Subnational estimates at NAICS3 level
    - 445 – Food and Beverage Stores
- Produces national, regional, divisional, and geographic estimates
  - Geographic estimates = state or an assembly of states

United States®
Census
Bureau

Annual Economic Surveys (2022)

Survey	Sampling Unit	Estimation Levels	
Annual R...	Company and EIN	National:	2-through-6-digit NAICS
Annu...	Company and EIN	National:	2-through-5-digit NAICS by Type of Operation
Se...	Company and EIN	National:	2-through-6-digit NAICS by Tax Exempt Status
A...	Company	National:	3- or 4-digit NAICS
Re...	Company		
M3 ...	Company	National:	M3 industry category
Annual S...	Establishment	National:	2-through-6-digit NAICS
		Subnational:	2-through 4-digit NAICS (50 states plus DC)

Annual Integrated Economic Survey

United States® Census Bureau

Sectors sampled

	Sector
Manufacturing	31-33
Wholesale Trade	42
Retail Trade	44-45
Transportation & Warehousing	48-49
Information	51
Finance & Insurance	52
Real Estate & Rental & Leasing	53
Professional, Scientific, & Technical	54
Management of Companies & Enterprises Excluded	55
Administrative & Support, & Waste Management & Remediation	56
Educational	61
Health Care & Social Assistance	62
Arts, Entertainment, & Recreation	71
Accommodations & Food Services	72
Other Services (except Public Administration)	81

Research Objective

- Poststratification to correct for the fact that we were not able to sample at the state or industry level

    - Adjust national estimates to frame national total

    - Adjust geographic estimates to geographic frame total

- Conduct empirical research to develop a recommendation for weight adjustment procedures to ensure AIES tabulations additivity requirements are met

Additivity Requirements

National Industry Estimates

sample-based national industry estimates of measure of size (MOS) = national frame totals

Geographic Estimates

sample-based geographic industry estimates of MOS = geographic frame totals

sample-based division industry estimates of MOS = divisional frame totals

sample-based region industry estimates of MOS = regional frame totals

National Level Summations (NAICS3)

NAICS3 aggregated national industry estimates = aggregated geographic estimates

Weight Adjustment Considerations

- HT estimator for a population total $\hat{Y} = \sum_{i=1}^{n} d_i y_i$
  - $d_i = 1/\pi_i$ is the inverse of the inclusion probability of company i
  - d_i is the design weight as it is directly obtained from the sampling design

- Using known auxiliary information, such as population totals (T), modify the basic design weights such that

$$\hat{T} = \sum_{i=1}^{n} \boldsymbol{w_i} y_i$$

  is equal to T.
  - $\boldsymbol{w_i}$ is the adjusted weight that satisfies the equation.

AIES Considerations

- Have three administrative variables (payroll, employment, and receipt)
  - Payroll is used as the measure of size (MOS) in the sample design
  - Missing data for the other two

- All adjusted weights are non-negative

%SurveyCalibrate: Methods Tested

Method	Pro(s)	Con(s)
Linear	• Runs quickly • No errors	• Negative weights
Exponential	• No negative weights	• Seldom runs without errors in optimization • Slow to run
Logistic		• Could not get to run • Errors in optimization
Truncated Linear	• No negative weights	• Seldom runs without errors in optimization • Slow to run • Weights less than 1

Ratio Adjustment

This ratio adjustment method modifies the company sample design weights, d_i, such that $w_i = R \cdot d_i$ where the ratio, R, is the population (auxiliary) total divided by the sample total

$$\hat{R}_h^{s(NC)} = \frac{\sum_{g \in s} \sum_{i \in s} \sum_{j \in i} x_{hgij}}{\sum_{g \in s} \sum_{i \in s} \sum_{j \in i} w_i x_{hgij}}$$

ratio of the noncertainty frame MOS total to the noncertainty sample estimates of the MOS total within s (s is the domain k for national estimates and s is the domain for subnational estimates)

Simulation

- 9 hand-selected industries at the NAICS3 level
  - Less than 3,000 companies in the sampling stratum
  - 5,000 random samples per industry

- Horvitz-Thompson estimates compared to ratio estimates

- $RB = \dfrac{\left[\frac{1}{5000}\sum_i\left(\hat{Y}_{l,i}^{m} - Y_l\right)\right]}{Y_l}$

  - l is the industry estimate (NAICS6 or NAICS3 by geography)
  - m is the estimation method (HT or ratio)
  - i is the simulated sample
  - Y_l is the true value

Findings

- Ratio estimates had low bias levels
- Comparable to Horvitz-Thompson estimator
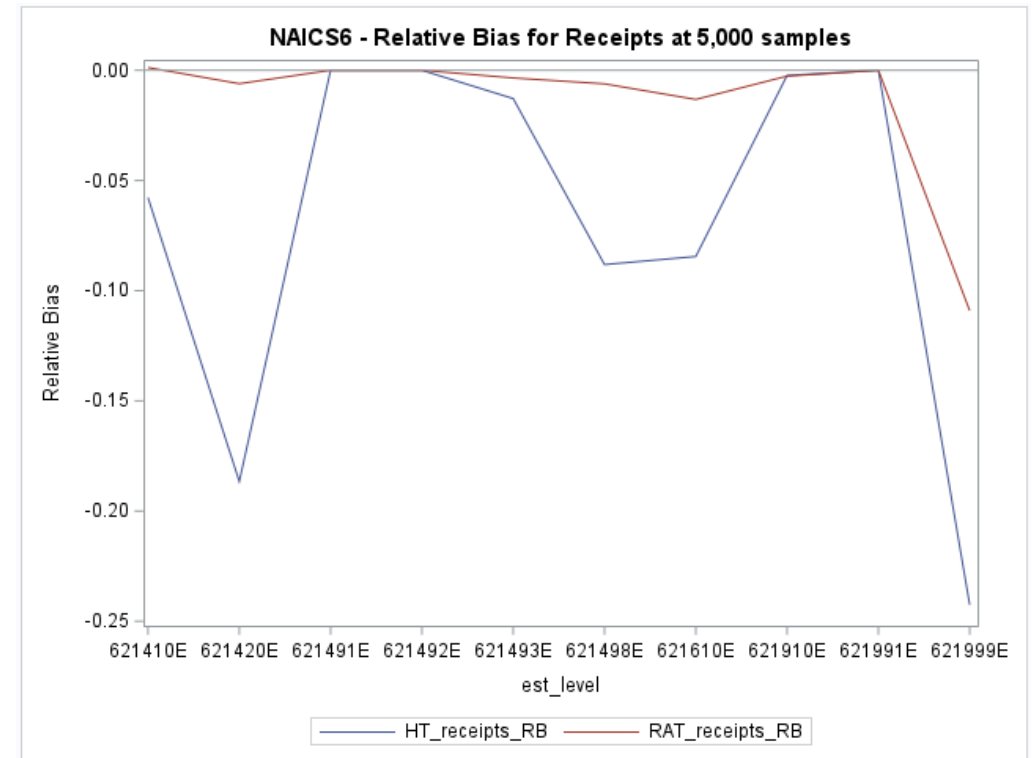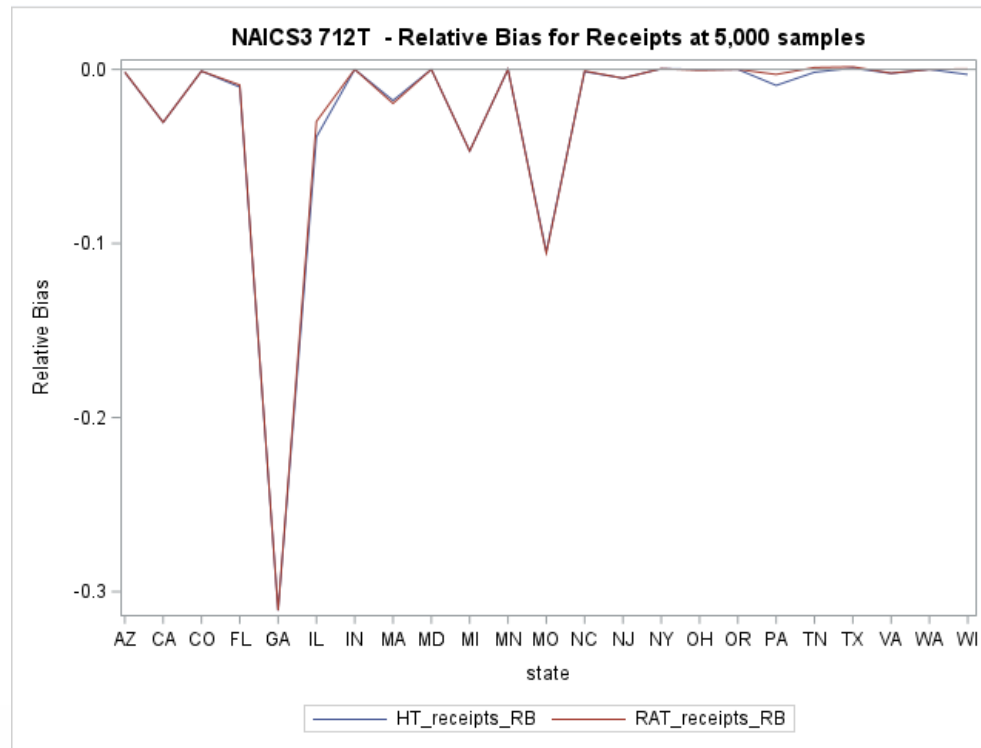
Next Steps

- Include additional auxiliary variables

- Repeat the research
  - Revisit after at least two years in production
  - Include additional industries

United States® Census Bureau

Questions, suggestions, comments?

quatracia.lucky@census.gov

yeng.xiong@census.gov