Using Published Margins of Error While Aggregating CTPP Tables and For Sensitivity Analysis

Jianzhu Li
Tom Krenzke

May 12, 2017

Disclaimer

The NCHRP Project 8-36C, Task 135 is conducted under contract to RAND and the final report is under review by the Transportation Research Board of the National Academy of Sciences, likely to be published this summer.

Outline

- Introduction

- Computing margins of error (MOEs) for aggregate estimates

- Creating replicated tables for the purpose of using MOEs to measure precision in analyses

- Summary and conclusion

Westat®

Introduction

- Key uses of ACS and CTPP data
    - The Census Transportation Planning Products (CTPP) comprise a set of special tabulations
        - Involve dozens of variables and provides millions of cell estimates and MOEs for small geographic units such as Traffic Analysis Zones (TAZs) that are roughly the size of Census blocks or block groups
        - Provide an important source of information for calibrating travel demand models
        - Facilitate transportation planning applications
    - The most recent CTPP is based on 2006-2010 American Community Survey (ACS) data

Westat

Introduction

- Address the challenges in using CTPP MOEs
  - TAZs were made to be small so that transportation planners could piece them together in different ways for planning purposes
  - Task 1: Solve the issues related to the aggregation of table cells and the computation of MOEs
    - Current practice treats subgroups as independent when estimating the MOE for combined subgroups
  - Task 2: Provide a method to incorporate the sampling error associated with CTPP estimates in travel demand modeling

Westat

Task 1 – Computing MOEs for Aggregate Estimates

- Traditional CTPP approach -- Naïve approach
  - Assume X_i's are independent
  - $X_{combined} = X_1 + X_2 + \cdots + X_n$

  - $MOE_{combined} = \sqrt{MOE_{X_1}^2 + MOE_{X_2}^2 + \cdots + MOE_{X_n}^2}$

  - Ignore covariance terms
  - Asiala (2012): resulting MOE is an overestimate (when aggregating) and seriously breaks down when aggregating more than four estimates

Westat

Task 1 – Computing MOEs for Aggregate Estimates

- Generalized Variance Function (GVF)
  - Let X = estimated total
  - Let V^2 = relative variance (sample variance divided by X^2)
  - A typical GVF is of the form… $V^2 = a + b/X$
  - Estimation:
    - Fitted GVF model using table estimates and MOEs sampled from all CTPP tables at TAZ and TAD level
    - Sampling was done within strata defined by types of tables and sizes of table estimates
  - Adjustment:

$$f_c = var_{gvf}(X_c)/var_{actual}(X_c), \ c = 1, .., C \ \text{(cells to combine)}$$

$$f_{uw} = \frac{1}{C}\sum_{c=1}^{C} f_c, \qquad f_w = \frac{\sum_{c=1}^{C} f_c * X_c}{\sum_{c=1}^{C} X_c}$$

Westat

Task 1 – Computing MOEs for Aggregate Estimates

- **Evaluation**
  - Selected a nationally representative sample of 200 TADs and their associated TAZs
    - Residence Table: Age (7) by MOT (3) for workers 16+
    - Workplace Table: Presence of children (2) by MOT (3)
    - Flow Table: Minority status (2)
  - Assumed TAZ level tables are aggregated to generate TAD level tables
  - CTPP TAD level tables were used as gold standard
  - Computed ratios of MOEs using CTPP MOE as base
  - Compared distribution of MOE ratios by number of TAZs combined

Westat®

Task 1 – Computing MOEs for Aggregate Estimates

- Evaluation Results: Age (7) by MOT (3)

| # of TAZs combined	# of TADs	MOE methods	5th	25th	Median	75th	95th	IQR
		Naive	0.91	0.99	1.04	1.12	2.97	0.13
		Trad. GVF	0.85	1.09	1.25	1.43	1.76	0.35
		U-adjusted GVF	0.76	0.90	0.99	1.08	1.24	0.17
<10	2,956	W-adjusted GVF	0.82	0.93	1.00	1.08	1.25	0.15
		Naive	0.94	1.04	1.15	1.31	1.95	0.28
		Trad. GVF	0.85	1.07	1.24	1.44	1.85	0.37
		U-adjusted GVF	0.76	0.89	0.98	1.08	1.28	0.20
10-49	2,480	W-adjusted GVF	0.82	0.92	1.00	1.10	1.29	0.17
		Naive	0.97	1.11	1.19	1.34	1.61	0.23
		Trad. GVF	0.87	1.01	1.16	1.32	1.65	0.32
		U-adjusted GVF	0.76	0.89	0.96	1.06	1.23	0.17
50-99	191	W-adjusted GVF	0.79	0.92	0.99	1.08	1.23	0.16
		Naive	1.00	1.18	1.29	1.48	1.83	0.31
		Trad. GVF	0.83	1.12	1.28	1.41	1.85	0.29
		U-adjusted GVF	0.70	0.87	0.95	1.07	1.22	0.20
>=100	87	W-adjusted GVF	0.73	0.90	0.98	1.09	1.28	0.20

Westat

Task 1 – Computing MOEs for Aggregate Estimates

- Evaluation Results: Age (7) by MOT (3)

# of TAZs combined	# of TAD estimates	MOE methods[1]	MOE Ratios ∈(0.8, 1.2)	MOE Ratios ∈(0.9, 1.1)
<10	2,956	Naive	0.82	0.67
		Trad. GVF	0.39	0.20
		U-adjusted GVF	0.84	0.55
		W-adjusted GVF	0.89	0.61
10-49	2,480	Naive	0.59	0.38
		Trad. GVF	0.42	0.22
		U-adjusted GVF	0.81	0.50
		W-adjusted GVF	0.86	0.56
50-99	191	Naive	0.55	0.22
		Trad. GVF	0.55	0.31
		U-adjusted GVF	0.82	0.55
		W-adjusted GVF	0.86	0.56
>=100	87	Naive	0.30	0.15
		Trad. GVF	0.33	0.14
		U-adjusted GVF	0.80	0.47
		W-adjusted GVF	0.78	0.52

Westat

Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- It is difficult to incorporate the sampling error associated with CTPP estimates in travel demand modeling

- Replicated tables approach
  - Given a CTPP table, generate several plausible sets of estimates that reflect the MOEs
  - Use each plausible table (or a subset for sensitivity analysis) in planned travel demand analysis to reflect the variation in the analysis
  - Results for each replicated table can be shown graphically

Westat®
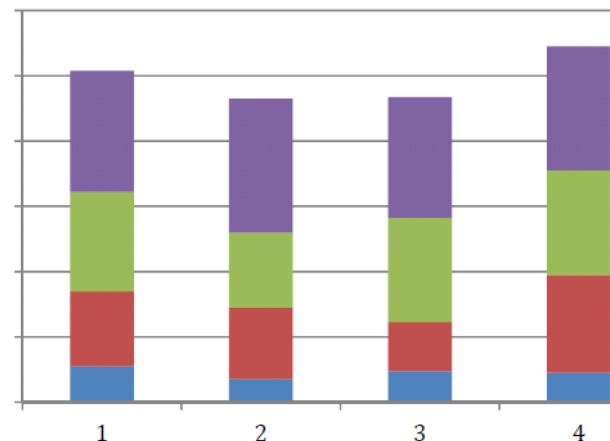
Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- Build models to take into account correlation among cell estimates

- Model assumptions for a given CTPP table with K cells

  - Assume a normal distribution model for overall weighted counts $X \sim N(\mu, \sigma^2)$

  - Assume a Dirichlet distribution model for cell proportions with parameters $(\alpha_1, \alpha_2, \ldots \alpha_k, \ldots, \alpha_K)$

    - $E(p_k) = \dfrac{\alpha_k}{\alpha_0}$

    - $Var(p_k) = \dfrac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0{}^2(\alpha_0 + 1)}$, where $\alpha_0 = \alpha_1 + \alpha_2 + \cdots + \alpha_K$

Westat

Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- Estimate model parameters
  - μ and σ^2 can be estimated from observed total $\hat{X}$ and its MOE
  - Estimate the parameters of the Dirichlet distribution from cell proportions
    - GVF method

$$p_k = \frac{\hat{X}_k}{\hat{X}}, \quad \hat{\alpha}_k = \left(\frac{\hat{X}}{\hat{b}} - 1\right) p_k$$

$$Var(p_k) = \frac{\hat{\alpha}_k(\hat{\alpha}_0 - \hat{\alpha}_k)}{\hat{\alpha}_0^2(\hat{\alpha}_0 + 1)} = \frac{\hat{b}}{\hat{X}} p_k(1 - p_k),$$

 - Distance Function Method

$$\text{Minimize } \sum_k \left(\frac{\hat{\alpha}_k(\hat{\alpha}_0 - \hat{\alpha}_k)}{\hat{\alpha}_0^2(\hat{\alpha}_0 + 1)} - v_k\right)^2$$

where $\quad v_k = var(p_k) = p_k^2 \left(\frac{var(\hat{X}_k)}{\hat{X}_k^2} - \frac{var(\hat{X})}{\hat{X}^2}\right)$ (see Wolter, 1985)

Westat

Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- Generate replicated tables in three steps
  - Randomly draw table total $\hat{X}$ from the normal distribution, multiple times
  - Randomly draw cell proportions p_k's from the Dirichlet distribution, multiple times
  - Derive cell counts $\hat{X}_k = \hat{X} p_k$ for each replicate

Westat®

Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- Census Bureau's "Variance Replicate Tables"
  - Published 80 variance replicate estimates for <u>select tables</u> from ACS 5-year
  - Can be used to calculate the MOEs of aggregated estimates
  - ACS uses a successive differences replication (SDR) variance estimation methodology
    - $variance = \frac{4}{80}\sum_{i=1}^{80}(Var\_Rep_i - ACS\ estiamte)^2$
  - Limitations
    - Does not cover all CTPP tables; has no plan to add more tables for CTPP
    - Graphing variance replicate estimates directly does not reflect the sampling error correctly

Westat

Conclusions and Summary

- Adjusted GVF method performed well in deriving MOEs for aggregated CTPP tables
- Replicated tables approach allows sampling error and perturbation error to propagate through to subsequent analysis and allows one to visually display results
- Toolkits
  - Computing aggregated MOEs
  - Generating replicated tables with graphical displays of the variation among the tables

Westat®

References

- Asiala, M. (2012). Topics on American Community Survey. Presented at the California Regional / Affiliate Data Center meeting, June 1, 2012.

- Wolter, K. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York, Inc.

Westat®