Westat

Improving lives through research

# Use of Published Margins of Error For Aggregating CTPP Tables and Sensitivity Analysis

*Jianzhu Li*
*Tom Krenzke*

November 13, 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.

# 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
- Demonstration of ToolKit

# Introduction

- Key uses of ACS and CTPP data
  - The Census Transportation Planning Products (CTPP) comprise a set of special tabulations
    - Provide estimates and MOEs for small geographic units such as Traffic Analysis Zones (TAZs)
    - 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 for combined cells
  - Task 2: Provide a method to reflect the sampling error associated with CTPP estimates in travel demand modeling

# 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

# Task 1 – Computing MOEs for Aggregate Estimates

- Generalized Variance Function (GVF)
  - A typical GVF is of the form… $V^2 = a + b/X$
    - X = estimated total
    - $V^2$ = relative variance (sample variance divided by $X^2$)
  - 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
    - X is between 0 and 100,000
  - 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 |
|---|---|---|---|---|---|---|---|---|
| <10 | 2,956 | 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 |
| | | W-adjusted GVF | 0.82 | 0.93 | 1.00 | 1.08 | 1.25 | 0.15 |
| 10-49 | 2,480 | 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 |
| | | W-adjusted GVF | 0.82 | 0.92 | 1.00 | 1.10 | 1.29 | 0.17 |
| 50-99 | 191 | 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 |
| | | W-adjusted GVF | 0.79 | 0.92 | 0.99 | 1.08 | 1.23 | 0.16 |
| >=100 | 87 | 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 |
| | | W-adjusted GVF | 0.73 | 0.90 | 0.98 | 1.09 | 1.28 | 0.20 |

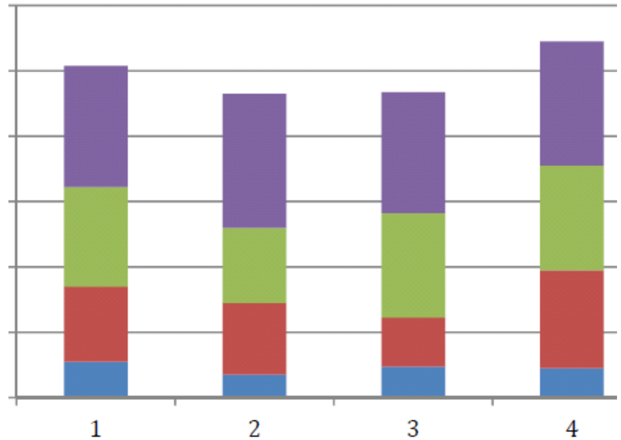
# Task 1 – Computing MOEs for Aggregate Estimates

- 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 |

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


Westat

# Task 2 – Creating Replicated Tables to Use MOEs in Analyses

- Estimate model parameters
  - $\mu$ and $\sigma^2$ 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 $v_k = var(p_k) = p_k{}^2 \left(\frac{var(\hat{X}_k)}{\hat{X}_k{}^2} - \frac{var(\hat{X})}{\hat{X}^2}\right)$ (see Wolter, 1985)

# 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 select tables from ACS 5-year
  - Can be used to calculate the MOEs of aggregated estimates
  - ACS uses a successive differences replication (SDR) variance estimation methodology
    - $variance = \frac{4}{80}\sum_{i=1}^{80}(Var\_Rep_i - ACS\ estimate)^2$
  - 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



Westat

# CTPP MOE ToolKit

- Excel worksheets and tutorial
- Functions
  - Estimate MOE for totals and proportions for combined subgroups (“CombineSubgroups” worksheet);
  - Compare proportions between two subgroups (“CompareSubgroups” worksheet); and
  - Replicate tables to reflect the published MOEs, for use in subsequent sensitivity travel demand analysis results
    - R macro


Westat

# Combine Subgroups

Click to open tutorial

CTPP Margin of Error Toolkit

Tutorial

Estimating MOE for Combined Subgroups

We emphasize that the Toolkit's GVF-based approaches should be used only when the aggregated total (*X*) is between 0 and 100,000, which is within the range of the mo

Generalized Variance Function parameters

| | |
|---|---|
| a | -0.00023 |
| b | 24.8988 |

GVF parameters

User Inputs

To estimate the proportion (*p*) for the combined estimate, enter the denominator *Y* value and its MOE:

| *Y* | *MOE* |
|---|---|
| 20300 | 914 |

Enter estimates to be aggregated

| *X* | MOE |
|---|---|
| 895 | 220 |
| 610 | 158 |
| 285 | 82 |
| 520 | 120 |
| 405 | 115 |
| 265 | 117 |
| 415 | 154 |
| 780 | 155 |
| 470 | 122 |
| | |

Input data

Results

Outputs

| Aggregated estimate | MOE(*x*) | | | |
|---|---|---|---|---|
| *X* | Naïve | GVF | GVF- uw adj | GVF- w adj |
| 4645 | 429.0 | 547.3 | 398.6 | 403.9 |

| | MOE(*p*) | | | |
|---|---|---|---|---|
| *p* | Naïve | GVF | GVF- uw adj | GVF- w adj |
| 0.2288177 | 0.01845 | 0.02682 | 0.01672 | 0.01702 |



Westat

# Compare Subgroups

**CTPP Margin of Error Toolkit**

**Comparing Two Proportions Between Subgroups**

**User Inputs**

Subgroup 1

| | |
|---|---|
| Proportion = | 0.04752 |
| MOE = | 0.04132 |

**Input data: two subgroup proportions for comparison**

Subgroup 2

| | |
|---|---|
| Proportion = | 0.03842 |
| MOE = | 0.00764 |

Enter I for independent subgroups or D for dependent subgroups:

| |
|---|
| I |

**Assumption on independence of subgroups**

**Outputs**

| | |
|---|---|
| Difference = | 0.01 |
| MOE = | 0.026 |
| Significant? | No |

**Test results**



Westat

# Create Replicated Tables

**CTPP Replicated Tables Toolkit**

| | X | | MOE.X |
|---|---|---|---|
| Total | 2985 | | 463 |
| | Xk | Pk | MOE.Xk |
| Cells 1 | 320 | 10.72% | 267 |
| 2 | 45 | 1.51% | 66 |
| 3 | 475 | 15.91% | 156 |
| 4 | 50 | 1.68% | 45 |
| 5 | 20 | 0.67% | 32 |
| 6 | 575 | 19.26% | 243 |
| 7 | 95 | 3.18% | 93 |
| 8 | 375 | 12.56% | 204 |
| 9 | 25 | 0.84% | 38 |
| 10 | 15 | 0.50% | 24 |
| 11 | 140 | 4.69% | 130 |
| 12 | 225 | 7.54% | 137 |
| 13 | 15 | 0.50% | 117 |
| 14 | 435 | 14.57% | 183 |
| 15 | 100 | 3.35% | 79 |
| 16 | 60 | 2.01% | 56 |
| 17 | 15 | 0.50% | 25 |

| Number of Replicated Tables |
|---|
| 5 |

Run

Input data



Westat

# Create Replicated Tables

| | | Rep1 | Rep2 | Rep3 | Rep4 | Rep5 |
|---|---|---|---|---|---|---|
| Total | | 0.9112 | 0.0319 | 0.4590 | 0.0171 | 0.9613 |
| Cells | 1 | 0.4739 | 0.2026 | 0.5913 | 0.1116 | 0.0055 |
| | 2 | 0.7115 | 0.3774 | 0.3263 | 0.3070 | 0.5701 |
| | 3 | 0.1590 | 0.2297 | 0.6635 | 0.7213 | 0.7222 |
| | 4 | 0.5229 | 0.2303 | 0.2129 | 0.1640 | 0.9157 |
| | 5 | 0.7493 | 0.4935 | 0.3192 | 0.7168 | 0.8231 |
| | 6 | 0.8032 | 0.0968 | 0.4240 | 0.6028 | 0.3245 |
| | 7 | 0.3790 | 0.5453 | 0.7757 | 0.7494 | 0.0872 |
| | 8 | 0.6268 | 0.1118 | 0.9095 | 0.5838 | 0.1862 |
| | 9 | 0.3958 | 0.6138 | 0.3791 | 0.1950 | 0.1652 |
| | 10 | 0.2391 | 0.6913 | 0.0453 | 0.2055 | 0.0876 |
| | 11 | 0.1957 | 0.1458 | 0.7587 | 0.0791 | 0.6799 |
| | 12 | 0.0538 | 0.4083 | 0.6697 | 0.4310 | 0.0119 |
| | 13 | 0.7331 | 0.6807 | 0.2208 | 0.3948 | 0.1270 |
| | 14 | 0.6218 | 0.8717 | 0.4034 | 0.6317 | 0.1362 |
| | 15 | 0.1131 | 0.7114 | 0.2854 | 0.9735 | 0.9828 |
| | 16 | 0.3545 | 0.6414 | 0.5373 | 0.4784 | 0.0077 |
| | 17 | 0.9221 | 0.4414 | 0.5254 | 0.3105 | 0.6758 |

Random Numbers

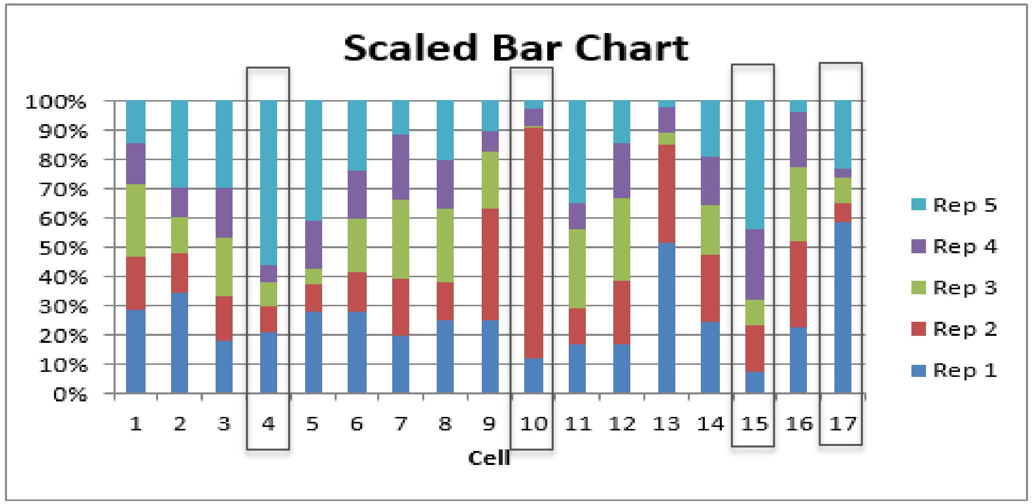
# Create Replicated Tables

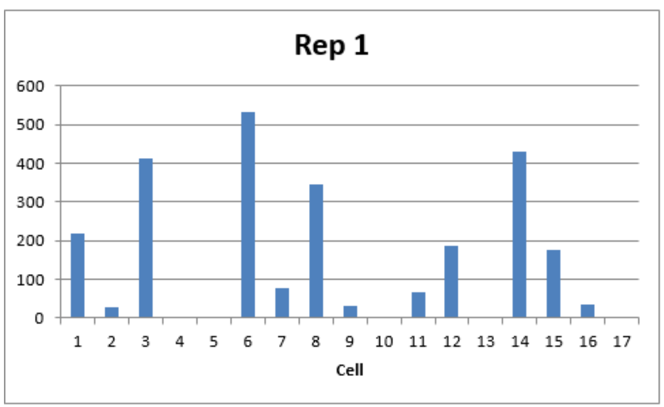
Replicated Tables - GVF Method

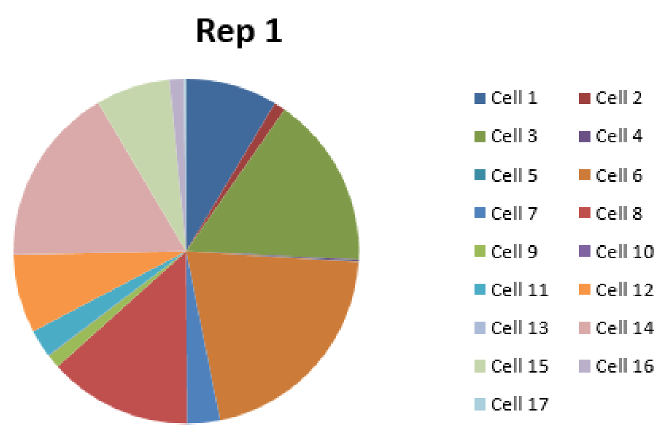
| | | Rep1 | Rep2 | Rep3 | Rep4 | Rep5 |
|---|---|---|---|---|---|---|
| Total | | 3364 | 2463 | 2956 | 2389 | 3482 |
| Cells | 1 | 371 | 234 | 322 | 176 | 190 |
| | 2 | 68 | 27 | 24 | 19 | 59 |
| | 3 | 444 | 376 | 497 | 433 | 736 |
| | 4 | 53 | 22 | 21 | 15 | 142 |
| | 5 | 33 | 12 | 6 | 20 | 49 |
| | 6 | 817 | 409 | 527 | 486 | 709 |
| | 7 | 89 | 89 | 123 | 99 | 53 |
| | 8 | 483 | 252 | 494 | 315 | 397 |
| | 9 | 15 | 23 | 12 | 4 | 6 |
| | 10 | 2 | 16 | 0 | 1 | 0 |
| | 11 | 107 | 77 | 170 | 54 | 221 |
| | 12 | 144 | 192 | 243 | 166 | 124 |
| | 13 | 23 | 15 | 2 | 4 | 1 |
| | 14 | 557 | 533 | 389 | 376 | 446 |
| | 15 | 55 | 116 | 65 | 177 | 321 |
| | 16 | 48 | 63 | 54 | 41 | 8 |
| | 17 | 54 | 6 | 8 | 3 | 21 |

# Create Replicated Tables

**Graphs - GVF Method**



# 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.