Calibration and Nonresponse Adjustment for the Quarterly Summary of State and Local Tax Revenues F-73 Survey

Justin Nguyen
Bac Tran

Program Research Branch
Governments Division
U.S. Census Bureau

United States Census Bureau
Economic Statistics

Disclaimer: *This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

Outline

- Overview of the Quarterly Summary of State and Local Tax Revenues (QTax)
  - Purpose
  - Primary Components
  - Data Products
- Methodology
  - Sample design
  - Estimation
  - Variance Estimation
- Estimation for Local Non-property Taxes (F-73)
- Next Steps

Overview of QTax

Purpose: To provide quarterly estimates of state and local government tax revenue at the national level and detailed tax revenue estimates for each state

Overview of QTax (cont'd)

There are three survey components in QTax:

- Local government property tax (F-71)
- State government taxes (F-72)
- Local government non-property taxes (F-73)

Overview of QTax (cont'd)

Data Products

- Table 1: National totals of state & local taxes
  - Estimates
  - Coefficients of Variation
  - Margins of Error
  - Total Quantity Response Rates (TQRR)
- Table 2: National totals of state government taxes
- Table 3: State government taxes by state

Overview of QTax (cont'd)

- F-71: Local Government Tax

T01: Property tax

- F-73: Local Government Non-property Tax

T40: Individual income

T41: Corporation net income

T09: General gross and receipts sales

T13: Motor fuel sales

T16: Tobacco product sales

T10: Alcoholic beverage sales

T24: Motor vehicle and operator's licenses

T15, T70, T71, T72: All other

Overview of QTax (cont'd)

- F-72: State Government Tax

T01: Property tax
T09: General Sales and Gross Receipts Taxes
T10: Alcoholic Beverages Sales Tax
T13: Motor Fuels Sales Tax
T16: Tobacco Products Sales Tax
T24: Motor Vehicles License
T25: Motor Vehicle Operators License
T40: Individual Income Taxes
T41: Corporation Net Income Taxes
T11: Amusements Sales Tax
T12: Insurance Premiums Sales Tax
T14: Pari-mutuels Sales Tax
T15: Public Utilities Sales Tax
T19: Other Selective Sales and Gross Receipts
T20: Alcoholic Beverages License
T21: Amusements License
T22: Corporations in General License
T23: Hunting and Fishing License
T27: Public Utilities License
T28: Occupation and Businesses License, NEC
T29: Other License Taxes
T50: Death and Gift Taxes
T51: Documentary and Stock Transfer Taxes
T53: Severance Taxes
T99: Taxes, NEC

Methodology: Sample Design for Local Government Property Tax (F-71)

- Past sample selected in 1997 and based on 1992 Census of Governments

- Sample redesigned in 2008 and current sample based on 2002 Census of Governments stratified on population size and tax revenues

Methodology: State Government Taxes (F-72)

- Census of all state governments
- National aggregations of detailed state government taxes

Methodology: Sample Design for Local Government Non-property Taxes

Local government non-property taxes (F-73)

- Non-probability sample design before 2010 quarter 4
- Since then, stratified simple random sample of local tax imposers based on the 2007 Census of Governments

Estimation: Local Government Property Tax (F-71)

- High response rate with growth rate imputation
- Horvitz-Thompson estimator for estimates and variance

Estimation: Local Government Non-property Taxes (F-71) (cont'd)

- Under ideal settings, there would be a 100% response rate and the Horvitz-Thompson (HT) estimator would be unbiased.

$$\hat{y}^{HT} = \sum_{i \in S} d_i y_i$$

where S is the sample and d_i is the survey design weight of the i^{th} unit.

Estimation: Local Government Non-property Taxes (F-73)

Non-response follow-up (NRFU)

A subsample of 600 units in the NR population of all quarters was selected

- The response rate (RR) of the NRFU sample was lower than expected
- Could not determine the nature of missingness mechanism.

Estimation: Local Government Non-property Taxes (F-73) (cont'd)

Missingness Mechanism

- Missing completely at random (MCAR)
  - Missingness is not related to the variables under study.
- Missing at random (MAR)
  - Missingness is related to the observed data but not the missing data.
- Missing not at random (MNAR)
  - Missingness is related to the variables under study

Assumption: F-73 is at least MAR

Estimation: Local Government Non-property Taxes Calibration Estimator (F-73) (cont'd)

- Adjust the survey weights so that the weighted sum of benchmark variables equals to a pre-determined set of population values.

$$\{d_i\}_{i\in S} \xrightarrow[NRAdjustment]{} \{w_i^*\}_{i\in S_R} \xrightarrow[Calibration]{} \{w_i\}_{i\in S_R}$$

where S_R is the responses subset of the sample S, such that

$$T_X = \sum_{i\in S_R} w_i x_i$$

where X is the known population total.

Estimation: Local government Non-property Taxes Calibration Estimator (F-73) (cont'd)

- If the response y is perfectly related to the variable X, $y_i = x'_i \beta$, $\forall i \in U$

  then $\hat{y}^{cal} = \sum_{i \in s} w_i y_i$ is an unbiased estimator of the population total Y .

- Generalized REGression (GREG) estimator is a simple form of calibration.

Estimation: Local Government Non-property Taxes Calibration Estimator (F-73) (cont'd)

- If not doing NR adjustment, i.e.,

$$\{d_i\}_{i\in S} \xrightarrow[Calibration]{} \{w_i\}_{i\in S_R}$$

the calibration performs on S_R, and assumes the non-respondents behave as the respondents.

- The adjusted weight for nonresponse

$$w^*_i = d_i / \theta_i$$

where $\theta_i = \Pr(i \in S_R | S)$ = response propensity estimated by a logit model.

A Logistic Response Model

- $\log\left(\frac{\theta_i}{1-\theta_i}\right) = {z_i}^T\gamma$

where z_i is a vector of covariates, in our case:

z_i = (type of governments, log(population size), log(tax2007), log(revenue2007), log(expenditure2007), log(debt2007), log(assets2007), number of times the unit responded in the whole year)

$$\theta_i = \frac{\exp({z_i}^T\gamma)}{1+\exp({z_i}^T\gamma)}$$

Coefficient vector γ is estimated from the data

Example: Local Government General Sales Tax (T09)

- Assumptions:
  - The total of sales tax in four quarters in the same year is approximately equal to the estimate from the Annual Finance Survey (AFS) of the same year for each local tax

$$y_q = T09_{2011q}\ ,\ q= 1,2,3,4$$

$$y = y_1 + y_2 + y_3 + y_4$$

- Approximately, we have

$$T09_{AFS,\ 2011} \approx y = y_1 + y_2 + y_3 + y_4$$

Example: Local Government General Sales Tax (T09) (cont'd)

- Calibration equations

$$\sum_{k \in S} w_k y_k = T_{AFS,2011}$$

$$\sum_{k \in S} w_k x_k = T_{AFS,2007}$$

Where w_k is calibration weight using linear distance, y_k is the sales tax in 2011, and x_k is the sales tax in 2007

Result: Local Government General Sales Tax (T09)

Calibration with NRFU Adjustment

$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$T_{AFS,2011}$
15,540,864 (CV=1.72%)	16,146,981 (CV=3.02%)	15,902,700 (CV=2.55%)	17,840,236 (CV=1.83%)	65,430,781

Variance Estimation

- $e_i = y - x'_i \beta$

$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$\hat{y}$
15,954,349	18,743,709	14,874,923	22,615,235	72,188,216

Pros & Cons

- Pros:
  - Almost unbiased
  - Valid if response propensity model is correct
  - Able to compute the variance for total
- Cons:
  - Does not work if the response y is not strongly related to the variable X
  - Invalid if the missingness is non-ignorable
  - Invalid if S_R is not a representative random sample

Next Steps

- Use more calibration variables
- Nonresponse bias study
- New sample design and questionnaire

Contact

Bac.Tran@census.gov

Justin.D.Nguyen@census.gov