Calibration Methods in the Quarterly Summary of State and Local Government Tax Revenues

Justin Nguyen, Bac Tran

U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-1912
Justin.d.nguyen@census.gov ; Bac.Tran@census.gov

Abstract

The Quarterly Summary of State and Local Government Tax Revenues (QTax) was designed to estimate quarterly property, sales, personal income, corporate income, and other taxes for state and local governments. The Bureau of Economic Analysis uses these estimates to develop estimates of the Gross Domestic Product (GDP). In this paper, we discuss the estimation approaches we developed and tested when the new survey sample for the local non-property component produced lower than expected response rates for several quarters and differences when reconciled with the reported Annual Survey of State and Local Government Finances data. A calibration model and non-response adjustment with statistical analyses were used to produce estimations of higher quality. In this paper, we describe our methodology and provide our results with an interpretation of the estimation using big data for the four QTax quarters in a given year when we attempted to model the non-response. The results provided equivalent and superior ability estimates when calibrated with non-response follow-up and thus, this estimation method with item calibration was recommended for the QTax when considering its lower than expected response rates. Furthermore, the validation response propensity model and variance estimation are discussed (using calibration weighting to adjust for non-response and coverage errors).

Key Words: Calibration, Horvitz-Thompson Estimator, Government Units

1. Introduction

The Quarterly Summary of State and Local Government Tax Revenues (QTax) is a compilation of three quarterly surveys conducted by the Governments Division (GOVS) of the U.S Census Bureau to estimate quarterly property, sales, personal income, corporate income, and other taxes for state and local governments. The Bureau of Economic Analysis (BEA) uses these estimates to develop estimates of the Gross Domestic Product (GDP). These quarterly surveys have been conducted continuously since 1962. The information contained in the summary provides information on a national basis for government tax collections.

Disclaimer: This report 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.

The QTax is comprised of three components: local government property taxes (F-71), state government taxes (F-72), and local government non-property taxes (F-73). The F-71 component is a stratified simple random sample, stratified by expected cost of collection, and is estimated with a Horvitz-Thompson (HT) estimator. The F-72 component is a census of all state governments. Prior to 2010, the F-73 component was a nonprobability sample, but was redesigned to a probability sample in the fourth quarter of 2010. In addition to the sample redesign, a new questionnaire was introduced to estimate the following taxes: income (individual and corporation net), sales (general and gross receipts, motor fuels, tobacco products, and alcoholic beverages), motor vehicles and operations licenses, and other taxes. This probability sample is a stratified simple random sample, with stratification based on state, type of government (county, city, township, and special district), and unit size.

This paper presents a validation of applying calibration to estimate national estimates for state and local governments. This calibration methodology has shown likely for less bias in estimating the national estimates for state and local governments.

2. Sample Design

2.1 Sub-stratification

The sample design for the F-73 is a two-stage stratified simple random sample. In the first stage, the strata are defined by state and type (counties, cities, townships, special districts, and school districts). After initial certainty units were selected, any units from sampling strata that contained less than six units were selected to be taken with certainty. If there are at least six units in a stratum, but no more than 10 units, then no further stratification is done. If there are more than 10 units in a stratum, but no more than 50 units, the stratum is marked to be divided into two sub-strata. Finally, if the stratum has over 50 units, the stratum will be divided into three sub-strata.

The Cumulative Square Root of the Frequency method is used to perform sub-stratification. For strata that are to be divided into two sub-strata, the algorithm is applied to the entire stratum with certainties removed. For strata that are to be divided into three sub-strata the approach is similar. After the first cutoff point is determined, the stratum defined by all the units that are greater than the cutoff point is divided once again using the same algorithm. This will give us two cutoff points, which define three sub-strata of the original stratum.

The measure of size used for the Cumulative Square Root of the Frequency method was determined by first summing all nine variables (general sales and gross receipts tax T09, individual income tax T40, corporate net income tax T41, motor fuels tax T13, tobacco product tax T16, alcohol beverage tax T10, motor vehicle and operator license tax T24, public utility tax T15, and all other taxes T70, T71 and T72) by state and type. Within the cell defined by each state and type combination, the variable with the largest total was used as the measure of size for all units in that cell.

2.2 Sample Selection

After all certainty units were removed and all sub-stratification was completed, the sample was selected using SAS ProcSurveySelect. A simple random sample (SRS) of five units was taken from each substratum. Sampling was conducted at the parent unit level. The final sample listing was at the individual unit level, where every child of a selected parent unit was included if the parent was selected. A unit is identified to be a child of second unit if (1) the first nine digits of the first unit's ID match the first nine digits of the second unit's ID, (2) the last five digits of the first unit's ID are not equal to '00000', and (3) the last five digits of the second unit's ID are equal to '00000'. Every unit with '00000' as its last five digits is considered a parent unit in this context, regardless of the presence of any children. These randomly selected units in addition to the certainty units make up the F-73 sample (see Table 1).

Table 1. Distribution of Units in the F-73 Universe by Sample

		Total
Sampled	Certainty	1,763
	Non-certainty	1,925
	Total	3,688
Not Sampled		31,050
Total		34,738

3. Estimation Methodology

In this paper, we focus on a specific kind of sales tax for the local governments, general and gross receipts (T09), individual income tax (T40), and corporate net income tax (T41), respectively. The Annual Survey of Local Government Finances (ASLGF) also collects T09, T40, and T41 information. The difference between these two surveys is the ASLGF is an annual survey, whereas the F-73 QTax is a quarterly survey. Due to the nature of these programs, the respondents of the ASLGF have more time to respond to the survey than those that respond to the QTax respondents. Additionally, the ASLGF includes an imputation process. Lastly, the ASLGF has a higher response rate than the quarterly survey. For these reasons, we used the ASLGF T09 total as one of the calibration variables in our model. Ideally, at the end of the survey year, the national total of four quarters of the T09 will be close to that of the ASLGF. One of the disadvantages of this estimation process for QTax is that the ASLGF data releases have a two-year lag (due to the availability of audited financial reports); therefore, QTax does not have the calibration total available at the necessary time in the production cycle to improve the calibration weight. In this paper, we used the available 2007 Census of Governments (CoG-F) data and the 2011 ASLGF survey year data as two calibration totals. We conducted our study with two different approaches: one with a non-response adjustment and the other one without a non-response adjustment.

3.1 Response Propensity Model

Each government unit in QTax is identified by state where it belongs to one of five types of government. Each unit also has auxiliary data provided in the 2007 ASLGF, e.g., revenue, expenditure, debt, assets, and annual sales tax, T09. Additionally, population estimates are available. We used these data as predictors for our response propensity model. The response indicator R was defined as if they responded in recent quarter (QTax quarter 1 of 2013) and 0 otherwise. We also introduced a variable (paradata), the number of times that a unit responded throughout the year of 2011. The proposed model is then defined as:

$$logit(R) = \beta' X \qquad (1), \text{ where}$$

$X = (\text{population size, revenue, expenditure, debt, assets, ALFIN annual tax T09, response count})$ and β is a slope vector. This model produced a best model fit statistic with R^2 of 0.86, and with 82 percent concordance.

3.2 Calibration

Calibration methods consist of reweighting units so that survey estimates coincide with known population totals or counts or percentages (also called benchmarks) from external sources. External sources include the census, administrative records or other available surveys. In our analysis, ASLGF is considered an external source as well as the 2007 Census of Governments: Finance.

The calibration estimator of a total is a linear estimator defined by

$$\hat{y}^{cal} = \sum_{i \in S} w_i(s) y_i \quad (2)$$

where the calibration weight $w_i(s)$ satisfied two constraints:

(a) $\sum_{i \in S} w_i(s)\boldsymbol{x}_i = X$ (calibration constraints)

(b) $w_i(s)$ are "close" to the design weight $d_i(s)$

The closeness of the calibrated weights to the original weights can be measured by a distance function $G(\frac{w_i}{d_i})$, where $G\left(\frac{w_i}{d_i}\right) \geq 0$, G(1)=0, differentiable with respect to w_i, the derivatives are continuous, and strictly convex. So, the total distance for the full sample is $\Sigma d_i G(\frac{w_i}{d_i})$. Minimizing the total subject to constraint (a) will yield a set of $w_i(s)$ that satisfy the above two conditions (Sarndal et al., 1992).

Due to the lower than expected response rate for QTax, the survey weights of responding units were adjusted to compensate for the nonresponse units. Weighting adjusted for nonresponse is finding $\{w_i^*, i \in S_R\}$ where S_R is the set of sample respondents. Then the weights $\{w_i^*\}$ will be calibrated to match the known totals. It is a two-step weighting system. A simple way to estimate w_i^* is to set $w_i^* = \frac{d_i}{p_i}$; where p_i is the response propensity of the ith unit (Sarndal et al., 2005). In our research p_i is estimated from the response model proposed in equation (1).

4. Results

We used the 2007 Census of Governments (CoG-F) as an external source to estimate the response propensity and to compare with the known total. We also used the secondary external source, 2011 ASLGF survey data, to provide the second total. With the additional information from the ASLGF data, the quality of national level estimates has been sufficiently reliable. This result yielded results which were similar to the results obtained when using the 2007 CoG-F. We ran four different quarters to estimate the non-property tax parameters for general sales and gross receipts tax (T09). The constraints for the known totals are then

(c) $\sum_{i \in S} w_i(s)\boldsymbol{T09}_{2007} = 61{,}076{,}440$ (T09's total in census 2007)

(d) $\sum_{i \in S} w_i(s)\boldsymbol{T09}_{2011} = 65{,}430{,}782$ (Estimate of T09's total in 2011)

where $\boldsymbol{T09}_{2007} = QTax\ T09_{2007}^{quarter\ 1} + QTax\ T09_{2007}^{quarter\ 2} + QTax\ T09_{2007}^{quarter\ 3} + QTax\ T09_{2007}^{quarter\ 4}$, and

$\boldsymbol{T09}_{2011} = QTax\ T09_{2011}^{quarter\ 1} + QTax\ T09_{2011}^{quarter\ 2} + QTax\ T09_{2011}^{quarter\ 3} + QTax\ T09_{2011}^{quarter\ 4}$.

Table 2, and Table 3 below show the results from calibration for four quarters of QTax in 2011 for the general sales and gross receipts tax (T09) and corporate net income tax (T41), respectively. As we proposed in our previous equations, the sum of the four QTax quarters are added up to the yearly total of the ASLGF data.

Table 2. Calibration estimates for general sales and gross receipts tax (T09)

QTax 2011 Estimates						**ASLGF 2011**
Nonresponse adjustment	Quarter 1	Quarter 2	Quarter 3	Quarter 4	**Total**	**ASLGF** 2011
No (Standard error)	14,965,291 (.0127)	16,487,296 (.0188)	15,537,350 (.0179)	18,440,964 (.0126)	65,430,901	65,430,781
Yes (Standard error)	14,879,196 (.0171)	15,998,818 (.0302)	16,005,647 (.0254)	18,547,120 (.0185)	65,430,781	

Table 3. Calibration estimates for corporate net income tax (T41)

QTax 2011 Estimates						**ASLGF 2011**
Nonresponse adjustment	Quarter 1	Quarter 2	Quarter 3	Quarter 4	**Total**	**ASLGF** 2011
No (Standard error)	1,562,973 (.0240)	2,831,021 (.0466)	1,451,004 (.0236)	1,318,772 (.0221)	7,163,770	7,163,771
Yes (Standard error)	1,562,973 (.0060)	2,831,021 (.0093)	1,451,004 (.0067)	1,318,772 (.0056)	7,163,770	

Another aspect that we considered in our study is the consistency in our estimated data using the calibration method. We revise our estimates for up to seven prior quarters, as the data are available. We have performed a statistical testing procedures required at a 90 percent level of significant. In all cases, calibration has shown the highest consistency for our data with smallest standard errors.

5. Future Research

There are a few remaining issues that need further research. First, in production we do not have real time ASLGF totals available, i.e., the known total in constraint (c) is not available at the time of the QTax estimation. To overcome this, we will conduct time series analysis research to project that total. Secondly, we estimate the variance of the residuals to approximate the estimated variance. We will continue research on the bootstrap variance or theoretical models to have better variances.

6. Conclusion

We validated the calibration method through a simulation and trend comparisons between the raw data that we received during the most recent QTax cycle. Although the response rates for QTax are lower than expected, calibration has shown a most reliable methodology for the estimation of QTax data. In all cases, calibration has produced smallest coefficient of variation (CV).

Calibration gives consistent estimates over the several rounds of data revisions that we normally produce and preserves the known totals. In addition, the response model helps to adjust for nonresponse units. Our recent research had shown that calibration outperforms the traditional Hortvitz-Thompson estimators about 80 percent of the time. This study leads to future research with an intensive simulation. Furthermore, the sample design weights are taken into consideration in this paper. With these features and investigations, we are able to produce better estimations for any future surveys.

References

Deville, J.-C and Sarndal, C.-E (1992). "Calibration Weighting", Journal of the American Statistical Association.

Kott, P. S. (2006). "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors", Survey Methodology, 32, 133-142.

Sarndal, C.-E. (2003). *Model Assisted Survey Sampling*, New York: Springer.

Sarndal, C.-E. and Lundstom, S. (2005). *Estimation in Surveys with Nonresponse*, Wiley.

Sarndal, C.-E. (2007). "The Calibration Approach in Survey Theory and Practice", Survey Methodology, 33, 99-119.