Using Paradata to Calibrate the Quarterly Summary of State and Local Government Tax Revenue

[1]Courtney Hill, [1]Justin Nguyen
[1]U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-1912
Courtney.Hill@census.gov ; Justin.D.Nguyen@census.gov

Abstract
In planning the estimation methodology for the Quarterly Summary of State and Local Government Tax Revenue (QTax), we decided to use additional data captured during the process of data collection, paradata, to improve the estimates. QTax is comprised of three components: local property tax, state tax, and local non-property tax. In this paper, we focus on local non-property taxes. These taxes include Individual Income, Corporation Net Income, and General Sales and Gross Receipts. We have paradata information on which governments responded each quarter. In our research, we used the paradata to improve sample design and the estimates. From the paradata, we developed a response propensity model to adjust the survey weights due to nonresponse. In this paper, we discuss how to use the paradata in our models and calibration estimators with adjusted weights and census calibration totals to produce the estimates that agree with the totals from the Annual Survey of State and Local Government Finances and from the Census of Governments.

Key Words: Calibration, Paradata, Quarterly Tax

1. Introduction

1.1 Survey Overview
The Quarterly Summary of State and Local Government Tax Revenue (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 three component surveys are local government property taxes (F-71), state government taxes (F-72), and local government non-property taxes (F-73). The Bureau of Economic Analysis (BEA) uses these estimates to develop estimates of the Gross Domestic Product (GDP). The summary contains information on a national basis for government tax collections. The F-71 component is a stratified simple random sample. It is stratified by county area population and a cost factor that depends on how many tax collection agencies are in a county. It is estimated with a Horvitz-Thompson (HT) estimator. The F-72 component is a census of all state governments. The F-73 component is a stratified probability sample design with an initial certainty stratum and a separate stratum for units whose response propensity can be estimated using a logistic regression model.

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.

In this paper, we present a validation of the calibration estimator in estimating the state and local government tax totals when using the paradata. This paper focuses on the F-73 component. Our research shows that the calibration estimates (1) align very well with the Annual Survey of Local Government Finances (ALFIN) tax totals, and (2) are less biased than those of the HT, assuming that the calibration ALFIN tax totals are reliable.

1.2 F-73 Questionnaires and Sample Designs

The F-73 component was redesigned to a probability sample in the fourth quarter of 2010 (2010Q4). Prior to this quarter, it was a non-probability sample. A new questionnaire was introduced for this redesign to estimate the following taxes in 2010Q4: individual income, corporation net income, general and gross receipts sales, motor fuels sales, tobacco products sales, alcoholic beverages sales, motor vehicles and operations licenses, public utility tax, and other taxes. This F-73 design is a stratified simple random sample, with stratification based on state and type of the governments (county, city, township, special district, and school districts). The survey experienced a unit response rate that was lower than the Census Bureau's standard; therefore, starting in the third quarter of 2013, a new questionnaire was designed to measure local general sales and gross receipts tax (T09), individual income tax (T40), and corporation net income tax (T41), see Dumbacher and Hogue (forthcoming 2014). The scope of the new F-73 component of QTax was reduced from 11 taxes to three to reduce respondent and processing burden, increase the response rate, and increase data quality.

In brief, the new F-73 sample is a stratified πps (Särndal et al., 1992) sample with (1) initial certainty criteria, (2) a sample from a separate stratum for units whose response propensity can be estimated using the paradata, and (3) a stratified πps sample from the units whose response propensity is undefined. The sample size is about 1,800 local government units.

1.3 Nonresponse and Paradata

We improved the F-73 estimates by adjusting the survey weight by using the paradata to account for the nonresponses, and then using those adjusted weights in the calibration estimation process. Paradata can be defined as information or data about the data collection process (Bethlehem, Cobben, & Schouten, 2011). An example of paradata is the number of times a respondent responds or changes their answers to a particular item of the questionnaire. The adjusted weight is determined by the response behaviors, response propensity, that is estimated by a logistic regression model in which paradata are used.

1.4 Outline

In this paper, the data used in our analysis are from quarter 1, 2011 (2011Q1) to quarter 4, 2013 (2013Q4). This paper presents how we used the paradata in estimation with nonresponse propensity adjustment. In Section 2, we discuss the sample designs. Section 3 describes the method of estimation, and Section 4 discusses the variance estimation. In addition to using the paradata to adjust the survey weight, a calibration estimator enables the agreement of the four quarter totals from a particular year of the QTax to the Annual Survey of Local Government Finances (ALFIN) total. This will be described in Section 5. Sections 6 and 7 discuss future research and conclude this paper, respectively.

2. Sample Design

2.1 Overview

The data included in this paper are from 2011Q1 to 2013Q4 with two different sample designs. The first sample design of F-73 QTax was for the 2010Q4 to the 2013Q2. This sample was designed to estimate the totals of nine tax variables at the national level. The new sample design of F-73 QTax is from the third quarter of 2013 to the present. This sample design contained only general sales and gross receipts tax (T09), individual income tax (T40), and corporation net income tax (T41). The description of these designs follows.

2.2 Sample Design in the fourth quarter of 2010 (2010Q4)

In short, the sample design for the F-73 component of QTax is a two-stage stratified simple random sample. In the first stage, the strata are defined by state and type. In this design, initial certainty units were identified by certainty criteria. Also, if a sampling stratum contained few units then it was taken with certainty. Otherwise, a simple random sample was used to select sample units within each stratum with some other conditions based on survey specifics. The sample size of this design is 3,688 units.

2.3 Sample Design in the third quarter of 2013 (2013Q3)

In brief, the sample design for the new F-73 is a stratified πps (Särndal et al., 1992) with initial certainty criteria and a separate stratum for units currently in the 2010Q4 sample whose response propensity can be estimated. A logistic regression model with the paradata as a covariate is used for the response propensity model. In this separate stratum where an estimated response propensity occurs, units above a certain cutoff are taken with certainty, and a πps sample is selected from the remaining units (Dumbacher and Hogue, forthcoming 2014). The measure of size is the sum of T09, T40, and T41. The sample size of this design is 1,821 units.

Table 1 shows the sample distribution of both F73 sample designs. It groups the units by sampled, not sampled, and the total. This total is the number of units in the frame from which the samples were drawn. Within the sampled group, it shows certainty, non-certainty, and the total number of units. This table helps show the large differences in the two samples when looking at the number of units contained.

Table 1: Sample Distribution for 2013 Quarterly Survey of Non-property Taxes

Unit Type		2010Q4 Total	2013Q3 Total
Sampled	Certainty	1,763	956
	Non-Certainty	1,925	865
	Total	3,688	1,821
Not Sampled		31,050	12,194
Total		34,738	14,015

Source: 2013 Quarterly Survey of Non-property Taxes

3. Estimation Methodology

3.1 Overview

In this paper, we focus on general sales and gross receipts tax (T09) for the local governments. However, individual income tax (T40) is also a variable that we tested as shown in the Results section. We used the calibration estimator to estimate those tax totals. We used the Annual Survey of Local Government Finances (ALFIN) and the 2007 Census of Governments as reliable external sources to be calibration totals. That is because those two sources collect the same information as QTax, but with a larger sample size; annually and in the Census.

3.2 Response Propensity Model

Due to low response rates in QTax, we have to account for nonresponse by adjusting the survey weight by introducing a response propensity model. The model uses paradata and auxiliary data provided in the 2007 CoG-F as predictors, for example, revenue, expenditure, debt, assets, and annual sales tax. The response indicator R was defined as if they responded in each quarter and 0, otherwise. Therefore, we have a response indicator for each quarter that we estimate. We also introduced a variable (response count) which is our paradata. Response count is the number of times that a unit responded throughout the previous four quarters before the chosen quarter. The proposed model is then defined as:

$$logit(p_i) = \beta' \boldsymbol{X_i} \qquad (1),$$

where $\boldsymbol{X_i}$ is a vector of covariates: log(population size), log(revenue), log(expenditure), log(debt), log(assets), log(2007 CoG-F annual sales tax T09), type of government, and response count. The probability of response is p_i and β is a slope.

This model was applied for each quarter from 2011 to 2013. Table 2 shows the goodness of fit (R^2) of model (1) from 2011Q1 to 2013Q4.

Table 2: R^2 values for each quarter for general sales and gross receipts tax (T09)

Year	Quarter	R^2
2011	1	.3795
	2	.9563
	3	.9579
	4	.9853
2012	1	.7779
	2	.9870
	3	.9826
	4	.9775
2013	1	.9895
	2	.9927
	3	.9325
	4	.9676

Source: 2011-2013 Quarterly Survey of Non-property Taxes

Because we are using paradata which tell us how many times a unit responded in the four previous quarters, the first quarter of 2011 does not have any of this information. Therefore, the R^2 is low in 2011Q1. The remaining quarters have good R^2, which shows that model (1) fits the data well.

3.3 Calibration

Calibration methods consist of reweighting units so that survey estimates coincide with known population totals from external sources. External sources include the Census, administrative records, or other available surveys. In our analysis, the ALFIN and the 2007 CoG-F are used as external sources, in other words, calibration totals.

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

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

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 $d_i(s)$

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

The survey weights of responding units were adjusted to compensate for the nonresponse units. This weight adjustment for nonresponse finds $\{w_i^*, i \in S_R\}$ where S_R is the set of sample respondents. Then the weights $\{w_i^*\}$ will be calibrated to match the known totals. Since the survey weights are adjusted for nonresponse and during calibration, it is called 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 i^{th} unit (Särndal et al., 2005). In our research, p_i is estimated from the response model proposed in equation (1).

4. Variance Estimation

Previously, we used the variance estimation formula from Kott and Chang (2010) which is defined by

$$\hat{V}(\hat{y}^{cal}) \approx \sum_{i \in S} ({w_i}^2 - w_i){e_i}^2$$

where $e_i = y_i - {x_i}'\beta$. The w_i's are the calibration weights; x_i's are the benchmark variables; and y_i's are the observed values.

SUDAAN 11 provides the variance estimates for a calibration estimator (Research Triangle Institute 2012). Therefore, we used this approach to calculate variance estimates in this paper.

5. Results

We used the 2007 CoG-F data as a calibration total. We also used a secondary external source, the ALFIN, to provide the second known totals. We projected the calibration totals using 2009-2011 data for 2012 and 2013 using the ratio estimator. We have 12 different quarters, 2011Q1 to 2013Q4, to calibrate for the General Sales and Gross Receipts Tax (T09). The constraints for the known totals are

$$(c) \qquad \sum_{i \in S} w_i(s)\boldsymbol{T09}_{2007} = 61{,}076{,}441$$

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

$$(e) \qquad \sum_{i \in S} w_i(s)\boldsymbol{T09}_{2012} = 66{,}421{,}663$$

$$(f) \qquad \sum_{i \in S} w_i(s)\boldsymbol{T09}_{2013} = 67{,}427{,}551$$

where

$\mathbf{T09}_{2007}$ is from the 2007 Census of Governments: Finance (CoG-F),

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

$\boldsymbol{T09}_{2012} = QTax\boldsymbol{T09}_{2012}^{quarter\ 1} + QTax\boldsymbol{T09}_{2012}^{quarter\ 2} + QTax\boldsymbol{T09}_{2012}^{quarter\ 3} + QTax\boldsymbol{T09}_{2012}^{quarter\ 4}$, and

$\boldsymbol{T09}_{2013} = QTax\boldsymbol{T09}_{2013}^{quarter\ 1} + QTax\boldsymbol{T09}_{2013}^{quarter\ 2} + QTax\boldsymbol{T09}_{2013}^{quarter\ 3} + QTax\boldsymbol{T09}_{2013}^{quarter\ 4}$.

Table 3, Table 4, and Table 5 show the results from the calibration for all four quarters of QTax in 2011, 2012, and 2013 for the General Sales and Gross Receipts Tax (T09) with and without using a response propensity model. Using the same method, we also obtained the estimates for individual income tax (T40). Tables 6, 7, and 8 display the results for individual income tax (T40).

Table 3: Calibration Estimates for General Sales and Gross Receipts Tax (T09) for 2011

Response Propensity	QTax 2011 Estimates				
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Yes	15,155,993	16,929,593	15,937,763	17,407,433	65,430,782
(CV %)	(1.36)	(0.80)	(2.58)	(1.24)	
No	15,162,949	17,104,657	15,014,425	18,148,751	65,430,782
(CV %)	(0.86)	(0.43)	(1.29)	(0.83)	

Source: Quarterly Summary of State and Local Government Tax Revenue

Table 4: Calibration Estimates for General Sales and Gross Receipts Tax (T09) for 2012

Response Propensity	QTax 2012 Estimates				
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Yes	15,088,993	17,630,498	15,409,433	18,292,740	66,421,664
(CV %)	(0.57)	(1.41)	(0.92)	(1.68)	
No	14,896,914	17,938,130	15,102,710	18,515,246	66,453,000
(CV %)	(0.84)	(0.91)	(1.22)	(0.45)	

Source: Quarterly Summary of State and Local Government Tax Revenue

Table 5: Calibration Estimates for General Sales and Gross Receipts Tax (T09) for 2013

Response Propensity	QTax 2013 Estimates				
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Yes	15,498,334	16,580,466	18,093,353	17,255,397	67,427,550
(CV %)	(1.02)	(12.34)	(3.14)	(4.66)	
No	16,039,645	17,893,659	16,953,619	16,540,628	67,427,551
(CV %)	(0.98)	(0.65)	(1.41)	(1.15)	

Source: Quarterly Summary of State and Local Government Tax Revenue

Table 6: Calibration Estimates for Individual Income Tax (T40) for 2011

Response Propensity	QTax 2011 Estimates					ALFIN Total
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total	
Yes	6,271,871	7,054,355	5,500,058	6,802,510	25,628,794	25,628,794
(CV %)	(1.94)	(3.24)	(8.54)	(4.49)		
No	6,545,378	6,913,679	5,049,754	7,119,983	25,628,794	
(CV %)	(0.82)	(1.91)	(3.98)	(3.29)		

Source: Quarterly Summary of State and Local Government Tax Revenue

Table 7: Calibration Estimates for Individual Income Tax (T40) for 2012

Response Propensity	QTax 2012 Estimates					ALFIN Total
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total	
Yes	6,758,779	8,003,461	5,158,901	6,057,975	25,979,116	25,979,115
(CV %)	(0.73)	(4.55)	(5.82)	(3.33)		
No	6,806,297	8,276,051	4,958,613	5,938,155	25,979,116	
(CV %)	(0.58)	(3.03)	(1.60)	(2.26)		

Source: Quarterly Summary of State and Local Government Tax Revenue

Table 8: Calibration Estimates for Individual Income Tax (T40) for 2013

Response Propensity	QTax 2013 Estimates					ALFIN Total
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total	
Yes	7,083,596	7,624,261	5,018,880	6,607,488	26,334,225	26,334,225
(CV %)	(0.93)	(1.41)	(4.46)	(1.15)		
No	7,126,076	7,644,593	5,098,027	6,465,528	26,334,224	
(CV %)	(0.74)	(0.81)	(1.38)	(1.24)		

Source: Quarterly Summary of State and Local Government Tax Revenue

The sum of the four quarters for each year equals or is approximately equal to the total of the Annual Survey of Local Government Finances (ALFIN) totals for T09 and T40. This shows that our calibration constraint is met. Because we are using a response propensity model to adjust for nonresponse, it is understandable that the CVs for this model could be slightly larger than without it because we are using an estimate within the calibration estimation. We have performed statistical testing procedures required at a 90 percent level of significance.

6. Future Research

We will extend our future research by conducting an intensive use of paradata. This includes the use of respondent inputs that are not completely submitted into the system to improve our estimation models and to increase data quality. We aim to provide our respondents with the ability to submit their tax information online through an enhanced system that could keep track of paradata information to reduce respondent and processing burden and encourage the respondent to continue to respond to our survey. We will also conduct a nonresponse study to assess the bias of the estimates, if any.

7. Conclusion

We found that the use of a response propensity model with paradata and the calibration estimators satisfied our needs. The method was validated by an intensive simulation method (Dumbacher and Hogue, forthcoming 2014). The paradata improve the goodness of fit from which the survey weights were adjusted correctly. In turn, the calibration estimation uses the adjusted weight to align with external reliable totals.

References

Bethlehem, J., Cobben, F., and Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken, NJ: John Wiley & Sons, Inc.

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

Dumbacher, B. and Hogue, C. forthcoming 2014. Using Paradata to Design the Quarterly Tax Sample. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

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

Phillip S. Kott & Ted Chang (2010) Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse, Journal of the American Statistical Association, 105:491, 1265-1275.

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

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

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

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Research Triangle Institute (2012). *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Research Triangle Institute.