

Using Paradata to Design the Quarterly Tax Sample

Brian Dumbacher¹, Carma Hogue¹

¹U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Brian.Dumbacher@census.gov, Carma.Ray.Hogue@census.gov

Abstract

The Quarterly Survey of Selected Non-Property Taxes (F-73) is a sample survey conducted by the Governments Division of the U.S. Census Bureau that collects data on local government revenue from non-property taxes. During its 2009 sample design, F-73 experienced a response rate that was lower than the U.S. Census Bureau's standard. In 2013, the Governments Division redesigned the sample and narrowed the scope of the survey from 11 taxes to three to try to reduce respondent burden, increase the response rate, and increase data quality. In this paper, we describe how paradata, which are data about the data collection process, were used in part of the sample redesign to select units with a high estimated response propensity. We also describe evaluation simulations that were performed to validate the redesign and fine-tune parameters.

Key Words: Paradata; Sample design; Response propensity; Logistic regression

1. Introduction

1.1 Survey Overview

The Quarterly Survey of Selected Non-Property Taxes is a sample survey conducted by the Governments Division of the U.S. Census Bureau that collects data on local government revenue from non-property taxes. This survey is known as F-73, the code of the corresponding questionnaire (U.S. Census Bureau, 2013). F-73 provides quarterly estimates of national totals and is one of three components that make up the Quarterly Summary of State and Local Government Tax Revenue (QTax). The two other components, F-71 and F-72, collect data on local government property tax revenue and state government tax revenue, respectively. QTax data play an important role in developing the Gross Domestic Product, a key economic indicator produced by the Bureau of Economic Analysis. See U.S. Census Bureau (2014) for more information on F-73 and QTax in general. Also, see National Research Council (2007, pp. 133-134) and Graham and Zamperini (2011) for historical information.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical, methodological, technological, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

1.2 Response Rate

For F-73, local governments are mailed advance notices and paper questionnaires and have the option of responding via the internet. The National Processing Center and QTax analysts at the Census Bureau follow up with non-respondents via email and phone.

During the 11 quarters of its 2009 sample design, F-73 experienced a unit response rate lower than expected. Although QTax estimates could not be released for the local non-property tax component alone, the total quantity response rates for estimates of state and local government totals met Census Bureau standards. It is believed that the unit response rate for F-73 was low because the fourth quarter of 2010 (2010Q4), which was the first quarter of data collection under the 2009 design, marked the first time that estimates for F-73 were based on a probability sample. The small units in the survey had never been surveyed before, and local governments were not prepared to give tax revenue information every quarter. Figure 1 shows the unit response rate during the 2009 sample design. [Even though it went into effect in 2010Q4, the 2009 sample design is so named because it was researched in 2009.]

Because a six-quarter bridge study was conducted during the first six quarters of the survey (Hogue, Zamperini, & Lee, 2012), dual processing was done for the two sampling methodologies. This meant that non-response follow-up efforts were not as intense as they were after the bridge was over. Despite efforts to increase the response rate through non-response follow-up, the rate did not climb, and it became apparent that the scope of the survey had to be narrowed.

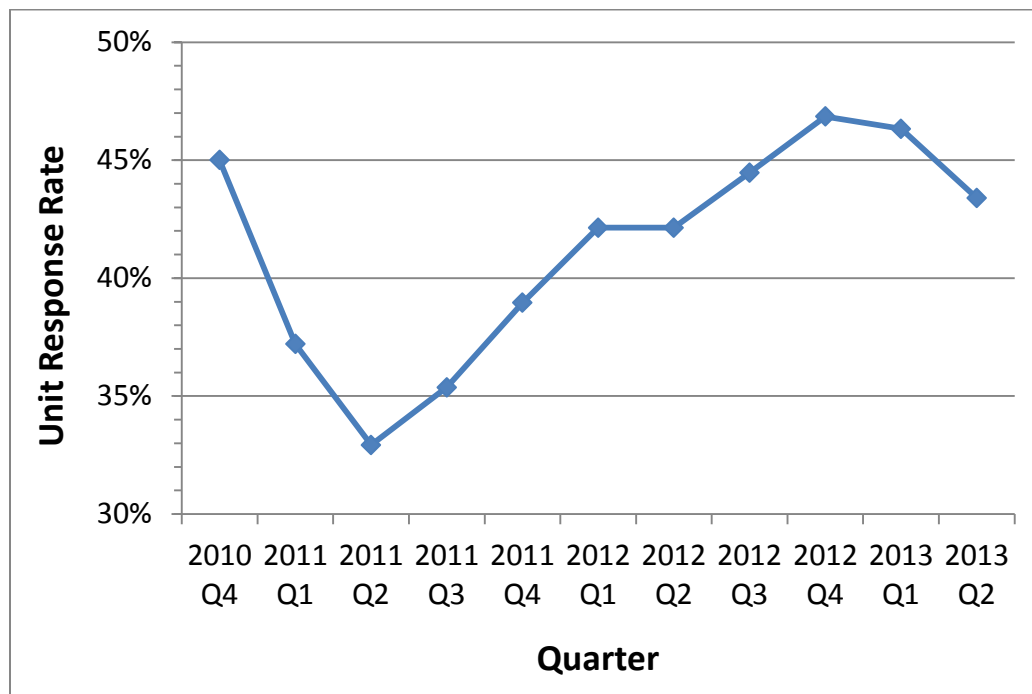


Figure 1: Unit Response Rate During the 2009 Sample Design. Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

1.3 Survey Redesign

In 2013Q3, the scope of F-73 was narrowed from 11 non-property taxes to three to try to reduce respondent burden, increase the response rate, and increase data quality. Table 1 lists the non-property taxes in scope during the 2009 design and the new 2013 design. The three taxes now in scope are general sales and gross receipts tax (T09), individual income tax (T40), and corporate net income tax (T41).

Table 1: Local Government Non-Property Taxes

Code	Tax	In Scope	
		2009 Design	2013 Design
T09	General sales and gross receipts	✓	✓
T10	Alcoholic beverages sales	✓	
T13	Motor fuels sales	✓	
T15	Public utilities sales	✓	
T16	Tobacco products sales	✓	
T19	Other sales and use	✓	
T24	Motor vehicles license	✓	
T25	Motor vehicle operators license	✓	
T40	Individual income	✓	✓
T41	Corporate net income	✓	✓
T99	All other	✓	

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

This fundamental change in the survey necessitated selecting a new sample from the universe of local governments that impose at least one of the three non-property taxes now in scope – T09, T40, and T41. This sample redesign offered an additional opportunity to increase the response rate by targeting governments that were in sample during the 2009 design and thought to have a high propensity of responding again if selected to be in the new sample.

1.4 Outline

The rest of this paper is organized as follows. Section 2 describes the paradata and auxiliary data used during sample selection. Section 3 describes the new sample design and the logistic regression model used to estimate response propensity. Section 4 describes evaluation simulations that were performed to validate the design and fine-tune parameters. Section 5 breaks down the selected sample by various characteristics. Finally, Sections 6 and 7 discuss limitations and ideas for future research, respectively.

2. Paradata and Auxiliary Data

2.1 Paradata

Paradata are data about the data collection process (Bethlehem, Cobben, & Schouten, 2011). They are available only for sample units and consist of information such as the number of contact attempts and response history. The paradata that we used for the sample redesign were F-73 response history data for 2011Q4 through 2013Q1. These six quarters were part of the 2009 sample design and were the most recent quarters for which data were available at the time of sample selection. Using these paradata, we constructed a variable called *response_history* that equals the number of times the unit responded during the five quarters 2011Q4 through 2012Q4. Figure 2 is a bar chart of the units in scope for the 2013 sample design broken down by *response_history*. Each bar is broken down further by response in 2013Q1. It shows that *response_history* is highly associated with response in 2013Q1, as expected. In other words, past response is a great indicator of future response. Figure 2 also shows that F-73 has many repeat respondents and repeat non-respondents.

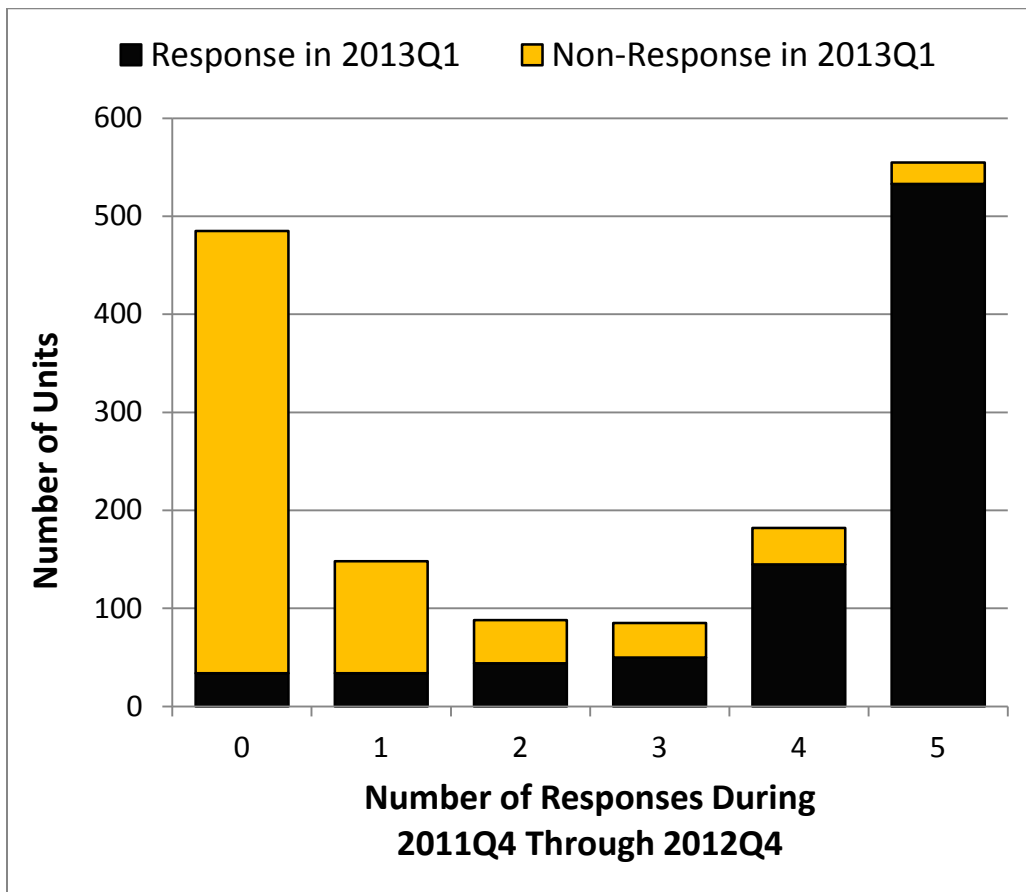


Figure 2: Response History. Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

2.2 Auxiliary Data

In years ending in “2” and “7,” the Census Bureau conducts the Census of Governments (CoG), which collects data on the operations, employment, and finances of the approximately 90,000 governments in the United States. A large amount of auxiliary information in the form of government-level CoG data was available to us. Assets, debt, revenue, expenditures, and 2007 values for T09, T40, and T41 came from the Finance component of the 2007 CoG. Also, because of overlap between F-73 and the intercensal Annual Survey of Local Government Finances, 2010 values for T09, T40, and T41 were available for many units.

2.3 Sampling Frame

The sampling frame for the 2013 sample design of F-73 is a list of local governments identified as imposing T09, T40, or T41. Subject matter experts provided a separate list for each tax, and then we merged them. To this frame were added the paradata and auxiliary data described above. Missing values for 2007 were filled in with 2010 values, if available. In the end, the final sampling frame consisted of 14,015 local governments. Table 2 breaks down the sampling frame by tax structure and shows that most governments impose only one of the three taxes.

Table 2: Sampling Frame by Tax Structure

Taxes Imposed	Frequency
T09	8,944
T40	4,678
T41	39
T09, T40	74
T09, T41	1
T40, T41	275
T09, T40, T41	4
Total	14,015

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

3. Sample Design

3.1 Overview

In short, the 2013 sample design for F-73 is stratified probability-proportional-to-size π ps (Särndal et al., 1992) with initial certainty criteria and a separate stratum for units currently in the 2009 sample whose response propensity can be estimated using a logistic regression model with *response_history* as a covariate. In this separate stratum, units with an estimated response propensity above a certain cutoff are taken with certainty, and a π ps sample is selected from the remaining units. It is described in Section 4 how the

cutoff and sample size were determined. Figure 3 illustrates the sample design by partitioning the universe into six parts. The parts are drawn roughly to scale, and the notation \hat{p} refers to the estimated response propensity. The dotted lines indicate samples of non-certainties.

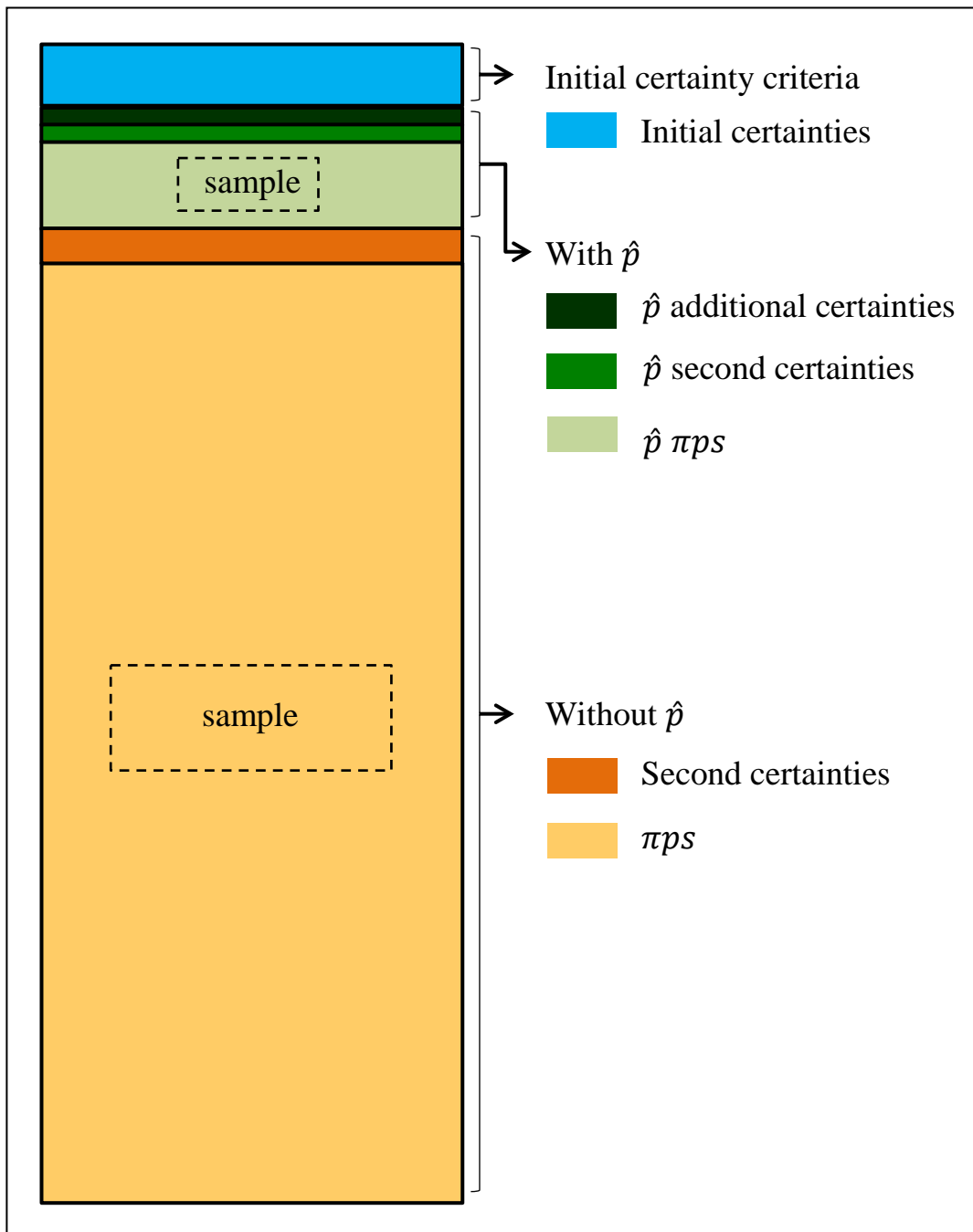


Figure 3: Illustration of the 2013 Sample Design

3.2 Requirements

The sample was designed to meet the following requirements. [The relative root mean squared error (RRMSE) of an estimator is the square root of the estimator's mean squared error divided by the value it is estimating. The RRMSE is similar to the coefficient of variation but takes bias into account.]

- The RRMSEs for national estimates of T09, T40, and T41 using the production calibration estimator should be less than three percent
- Because of cost constraints, the sample size should be around 1,800
- For consistency, the sample should contain many units from the 2009 design

The following subsections describe the sample design in more detail.

3.3 Initial Certainties

First, initial certainty criteria are applied. These are some of the same initial certainty criteria that were used for the 2009 design, so this provides some consistency in sample composition. The criteria are based on government type. The Census Bureau classifies governments into five types: counties, cities, townships, special districts, and independent school districts. Counties, cities, and townships tend to perform many functions, while special districts and independent school districts tend to perform one or a limited number of functions. Examples of special districts are transportation authorities, cemetery districts, and utilities. The initial certainty criteria are:

- Counties with population $\geq 100,000$
- Cities with population $\geq 75,000$
- Townships with population $\geq 50,000$
- Independent school districts with school enrollment $\geq 10,000$
- Special districts with the function
 - Electric power in Utah and Wisconsin
 - Gas supply in any state

3.4 Units Currently in Sample

Next, a stratum is created consisting of units currently in the 2009 sample. To these units we fit a logistic regression model predicting response in 2013Q1, the most recent quarter for which F-73 data were available at the time of sample selection. We take this estimated response propensity, which is denoted \hat{p} , as an indicator of future response. Table 3 describes the covariates and summarizes the fitted model. Some units currently in sample could not be used to fit the model because of missing values for some covariates.

Table 3: Summary of Logistic Regression Model

Variable	Description	$\hat{\beta}$	Wald χ^2	p-value
<i>intercept</i>	Intercept	-3.84	300.49	< 0.0001
<i>log_assets</i>	Natural logarithm of assets from 2007	-0.78	259.14	< 0.0001
<i>log_debt</i>	Natural logarithm of debt from 2007	0.14	23.42	< 0.0001
<i>log_revenue</i>	Natural logarithm of revenue from 2007	0.12	0.59	0.4436
<i>log_expenditures</i>	Natural logarithm of expenditures from 2007	0.48	10.28	0.0013
<i>government_type</i>	City indicator	0.23	5.16	0.0232
	Township indicator	0.51	3.33	0.0681
	Special district indicator	-1.42	14.28	0.0002
	Independent school district indicator	-0.68	23.85	< 0.0001
<i>log_tax</i>	Natural logarithm of sum of T09, T40, and T41 from the frame	0.19	22.04	< 0.0001
<i>response_history</i>	Number of times the unit responded in the period 2011Q4 through 2012Q4	0.96	2,799.50	< 0.0001
Model Fit Statistics				
Likelihood ratio = 5,211.25 (p-value < 0.0001)		Area under ROC curve = 0.93		

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

Units with \hat{p} greater than some cutoff are made \hat{p} additional certainties. From the units with \hat{p} less than the cutoff, we select a πps sample. Section 4 describes how the cutoff and sample size were chosen. The measure of size is the sum of T09, T40, and T41 from the sampling frame, and units with a measure of size equal to zero or missing are assigned a measure of size equal to the median measure of size of all the units below the cutoff. At this point, there are governments with a measure of size so large that they would have a probability of selection greater than one. To avoid this problem, we include these units in sample with certainty and call them \hat{p} second certainties. The original sample size is decreased as \hat{p} second certainties are determined.

3.5 Units Currently Not in Sample

The remaining units are those currently not in the 2009 sample or without an estimated response propensity. From these units we select a stratified πps sample where the strata are formed by state and type, and the measure of size is, again, the sum of T09, T40, and T41 from the sampling frame. Units with a measure of size equal to zero or missing are assigned a measure of size equal to the median measure of size in their stratum. As with

the \hat{p} πps sample, there are governments with a measure of size so large that they would have a probability of selection greater than one. To avoid this problem, we include these units in sample with certainty and call them second certainties. Table 4 reports the πps sample size n_h by stratum h . The notation N_h stands for the number of units in stratum h on the sampling frame. These sample sizes n_h are based on historical sample sizes and are modified to strike a balance between efficiency and adequate representation for small strata. We allocate more sample to certain strata in Indiana, Iowa, Kentucky, Ohio, and Pennsylvania because they contain many units that impose T40 and T41. These base πps sample sizes are decreased as second certainties are determined.

Table 4: Base πps Sample Sizes by Stratum

Stratum h	Sample Size n_h
151 – Indiana counties	14
152 – Indiana cities	
165 – Iowa school districts	
181 – Kentucky counties	
182 – Kentucky cities	
362 – Ohio cities	28
365 – Ohio school districts	
392 – Pennsylvania cities	
393 – Pennsylvania townships	
395 – Pennsylvania school districts	
$N_h \leq 4$	N_h
$5 \leq N_h \leq 8$	5
$9 \leq N_h \leq 40$	9
$N_h > 40$	13

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

4. Simulations

4.1 Overview

We performed a series of simulations to validate the design and determine values for some of the parameters, most notably the \hat{p} cutoff and the sample size for the units below the \hat{p} cutoff. Taking advantage of the external validation offered by five-year CoG, we used data from the Finance components of the 2002 and 2007 CoG and restricted attention to the governments appearing in both years. For each simulation, the 2002 data served as the sampling frame, from which we selected 200 independent samples according to the design under consideration. These samples were merged with the 2007 data and used to estimate the known 2007 totals. Estimates were calculated for T09 and T40 only. In the restricted universe, only two governments had values for T41, and both were large enough to be made certainties.

4.2 Horvitz-Thompson Simulation

In the first simulation, we used the Horvitz-Thompson estimator and considered different combinations of \hat{p} cutoff and sample size below the \hat{p} cutoff to find a combination resulting in an overall sample size around the desired 1,800 and to make sure we were meeting the RRMSE requirements. The stratum sample sizes for the regular πps part of the design were kept constant. We assumed a 100 percent response rate, which is unrealistic, but the point of this first simulation was to make sure the design worked under the best possible conditions. This was the first time using a πps design for F-73, and it was unclear how the separate \hat{p} stratum would affect efficiency. If the design was not meeting the RRMSE requirements, then we could make fundamental changes to the design early. Table 5 gives the results from this first simulation. All combinations resulted in RRMSEs below three percent, but the combination (\hat{p} cutoff = 0.95, sample size = 125) resulted in an overall sample size closest to the desired 1,800.

Table 5: RRMSEs and Sample Sizes From the Horvitz-Thompson Simulation

Sample Size Below \hat{p} Cutoff	\hat{p} Cutoff							
	0.80		0.85		0.90		0.95	
100	T09	0.67%	T09	0.76%	T09	0.85%	T09	1.06%
	T40	0.70%	T40	0.81%	T40	0.81%	T40	1.00%
	<i>n</i>	1,974	<i>n</i>	1,941	<i>n</i>	1,875	<i>n</i>	1,782
125	T09	0.59%	T09	1.02%	T09	0.93%	T09	0.90%
	T40	0.75%	T40	0.89%	T40	0.77%	T40	0.95%
	<i>n</i>	1,999	<i>n</i>	1,966	<i>n</i>	1,900	<i>n</i>	1,807
150	T09	0.83%	T09	0.88%	T09	0.59%	T09	2.50%
	T40	0.72%	T40	0.84%	T40	0.88%	T40	0.92%
	<i>n</i>	2,024	<i>n</i>	1,991	<i>n</i>	1,925	<i>n</i>	1,832
175	T09	0.63%	T09	0.87%	T09	0.77%	T09	0.70%
	T40	0.79%	T40	0.84%	T40	0.77%	T40	0.88%
	<i>n</i>	2,049	<i>n</i>	2,016	<i>n</i>	1,950	<i>n</i>	1,857
200	T09	0.59%	T09	2.05%	T09	0.74%	T09	0.61%
	T40	0.74%	T40	0.83%	T40	0.77%	T40	0.82%
	<i>n</i>	2,074	<i>n</i>	2,041	<i>n</i>	1,975	<i>n</i>	1,882

Source: U.S. Census Bureau, 2002 and 2007 Censuses of Governments: Finance, Quarterly Summary of State and Local Government Tax Revenue

4.3 Calibration Simulation

Next, focusing on the combination (\hat{p} cutoff = 0.95, sample size = 125), we considered the calibration estimator used in production and various unit response rates. To simulate a given unit response rate (URR), we did the following: before sample selection, the units were ordered from largest to smallest in terms of the sum of T09, T40, and T41, and the largest URR percent of them were made respondents. The remaining units were made non-respondents. This method of simulating response rates is based on the assumption that larger units have a higher propensity of responding. This assumption is supported by anecdotal evidence and the fitted logistic regression model summarized in Table 3.

Figure 4 plots the RRMSE versus simulated URR for T09 and T40. For simulated URRs in the realistic range 30 percent to 50 percent, the RRMSE requirements are met. For more information about the calibration estimator and current research in this area using paradata, see Hill and Nguyen (forthcoming 2014).

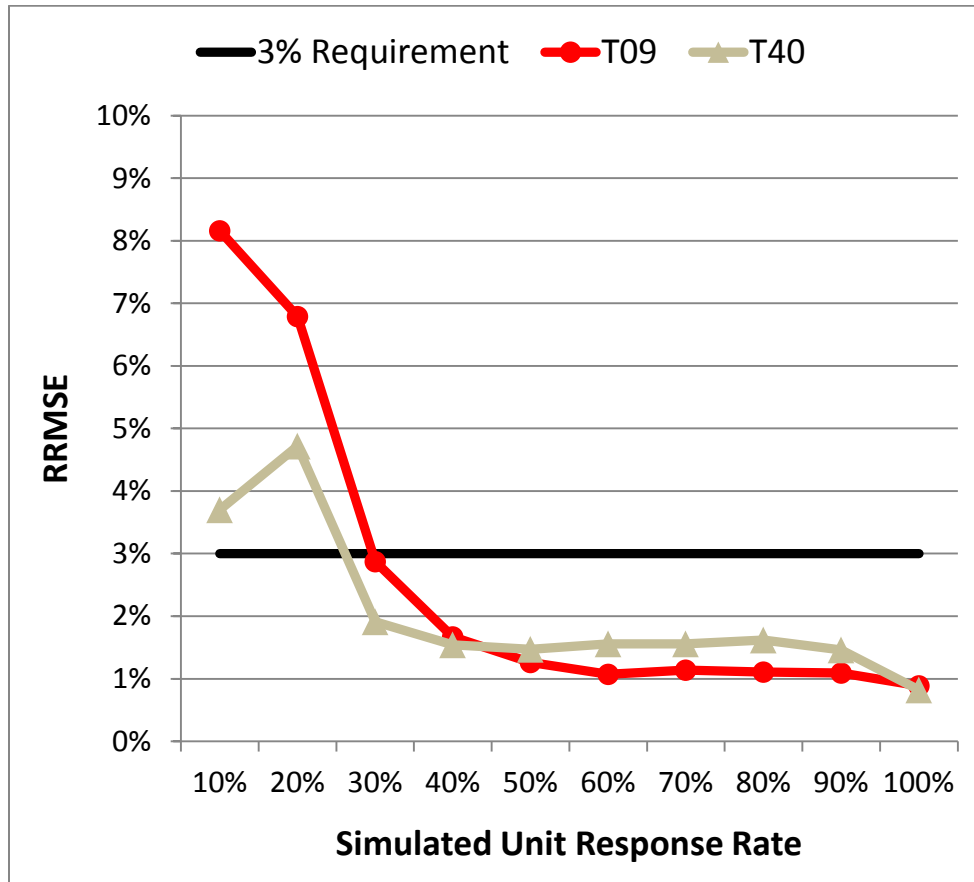


Figure 4: RRMSEs From the Calibration Simulation. Source: U.S. Census Bureau, 2002 and 2007 Censuses of Governments: Finance, Quarterly Summary of State and Local Government Tax Revenue

As an aside, we performed an additional calibration simulation after collecting data in 2013Q3, the first quarter of the new design. To improve the simulated URR method used during the short research period before sample selection, we estimated response rates separately for certainties and πps units and randomly determined respondents based on a Bernoulli experiment for each unit in sample. The 30 largest units in sample were made respondents. Based on this simulation, we estimated the RRMSEs for T09 and T40 to be 1.37 percent and 1.13 percent, respectively.

4.4 Final Parameter Values

After validating the design, we made small adjustments to the \hat{p} cutoff and the sample size below the \hat{p} cutoff in order to increase the sample sizes in certain strata and to increase representation for taxes T40 and T41. The calibration simulations were rerun to make sure the estimated RRMSEs did not change much. The final \hat{p} cutoff was set to 0.92, and the final sample size below the \hat{p} cutoff was set to 130.

5. Selected Sample

The selected sample contained 1,821 units. Tables 6 and 7 break down the sample and universe by unit type and by response in 2013Q1, respectively.

Table 6: Sample Breakdown by Unit Type

Unit Type		Sample	Percent of Sample	Universe	Sampling Rate
Initial certainty		645	35.4%	645	100.0%
\hat{p}	Additional certainty	140	7.7%	140	100.0%
	Second certainty	24	1.3%	24	100.0%
	<i>πps</i>	106	5.8%	578	18.3%
Second certainty		147	8.1%	147	100.0%
<i>πps</i>		759	41.7%	12,481	6.1%
Total		1,821	100.0%	14,015	–

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

Table 7: Sample Breakdown by Response in 2013Q1

Responded in 2013Q1?	Sample	Percent of Sample	Universe	Sampling Rate
Yes	610	33.5%	840	72.6%
No	341	18.7%	703	48.5%
Not in sample at the time	870	47.8%	12,472	7.0%
Total	1,821	100.0%	14,015	–

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

The first quarter of data collection under the new design was 2013Q3. Table 8 gives the URR for 2013Q3 by unit type. The overall URR is 56.3 percent, and the URR of 93.6 percent for the \hat{p} additional certainties is greater than the \hat{p} cutoff 0.92 determined using the logistic regression model. This is evidence that the survey redesign and the use of paradata in the sample design to target units helped improve the response rate, but it is

important to note that the effects of these two design changes are confounded. One limitation is that when multiple changes are made to a survey at once, it is difficult to separate the effects. On an added note, Table 8 reports the total sample size to be 1,820, one less than the total given in Tables 6 and 7. This is because one sample unit in Maryland was determined to be out of scope and was dropped from the sample.

Table 8: 2013Q3 Unit Response Rate by Unit Type

Unit Type		Sample	Unit Response Rate
Initial certainty		645	62.0%
\hat{p}	Additional certainty	140	93.6%
	Second certainty	24	79.2%
	<i>pps</i>	106	50.0%
Second certainty		146	61.6%
<i>pps</i>		759	43.6%
Total		1,820	—

Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue

6. Limitations

The time constraint for selecting the production sample limited our research. We only had one month to conduct research, perform evaluation simulations, and select the production sample, so we did not have sufficient time to conduct larger simulations and to explore response propensity models thoroughly. We found that the most important predictor of future response is response history, but other paradata such as the number of contact attempts would have been interesting to include in the analysis.

Data from the 2012 Census of Governments: Finance were not available. Instead, we used data from the Finance components of the 2002 and 2007 Censuses of Governments in our simulations.

7. Future Research

After data from the 2012 Census of Governments: Finance become available, we will select a new sample for F-73. At that time, we will conduct more sample design and paradata research. A couple ideas to explore include stratifying by response propensity and constructing a measure of size that is a function of response propensity. Similar simulations would be performed to examine the efficiency of the sample designs in combination with the calibration estimator used in production.

Also, anecdotal evidence and the fitted logistic regression model show that large units, i.e., units with large non-property tax revenue, tend to have a higher propensity of responding, but the relationship is not strong. Large units are perceived to have a large impact on the estimates and are followed up until they respond, so the number of contact attempts would be important to include in any analysis.

We should also revisit the initial certainty criteria and determine if they are really needed. Currently, one third of the units in the new 2013 sample are initial certainties. Efficiency might be improved by dropping the initial certainty criteria altogether and letting the πps sampling algorithm determine certainties based on the measure of size.

References

- Bethlehem, J., Cobben, F., and Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken, NJ: John Wiley & Sons, Inc.
- Graham, M., and Zamperini, B. (2011). Initiatives and Improvements to the Quarterly Summary of State and Local Government Tax Revenue. Governments Division Report Series, Research Report #2011-4. <www2.census.gov/govs/pubs/2011pubs/govsrr2011-04.pdf>
- Hill, C.M., and Nguyen, J. (forthcoming 2014). Using Paradata to Calibrate the Quarterly Summary of State and Local Tax Revenue. *2014 Proceedings of the American Statistical Association, Survey Research Methods Section*. Alexandria, VA: American Statistical Association.
- Hogue, C., Zamperini, B., and Lee, C. (2012). Bridge Report for the Introduction of the New Sample Design and Questionnaire for the Local Nonproperty Component of the Quarterly Summary of State and Local Tax Revenues. Governments Division Report Series, Research Report #2012-3. <www2.census.gov/govs/pubs/2012pubs/govsrr2012-03.pdf>
- National Research Council, Committee on National Statistics. (2007). *State and Local Government Statistics at a Crossroads*. Panel on Research and Development Priorities for the U.S. Census Bureau's State and Local Government Statistics Program. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York, NY: Springer-Verlag.
- U.S. Census Bureau. (2013). Quarterly Survey of Selected Non-Property Taxes: Form F-73. July 11, 2013. <www2.census.gov/govs/forms/2013/f73_2013.pdf>
- U.S. Census Bureau. (2014). Quarterly Summary of State & Local Taxes. <www.census.gov/govs/qtax>